

# Introduction to *Phenix* Refinement Tools for Low-Resolution Crystallography and Cryo-EM

Pavel Afonine

Lawrence Berkeley National Lab

November 10, 2016

# The *Phenix* project

Lawrence Berkeley National Lab

Paul Adams, Pavel Afonine,  
Nigel Moriarty, Billy Poon,  
Oleg Sobolev, Dorothee  
Liebschner



Los Alamos National Lab

Tom Terwilliger, Li-Wei Hung



Randy Read, Airlie McCoy,  
Tristan Croll, Rob Oeffner, Richard  
Mifsud

Cambridge University



Duke University

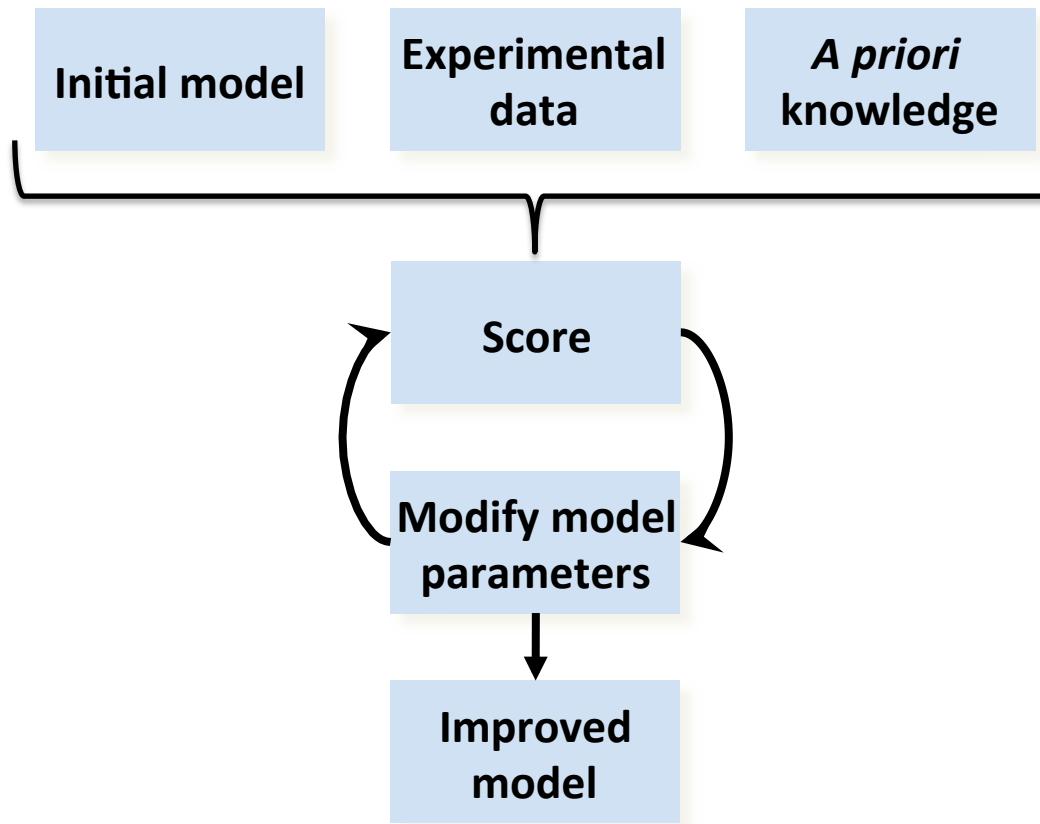
Jane & David Richardson,  
Chris Williams, Bradley Hintze



An NIH/NIGMS funded Program Project

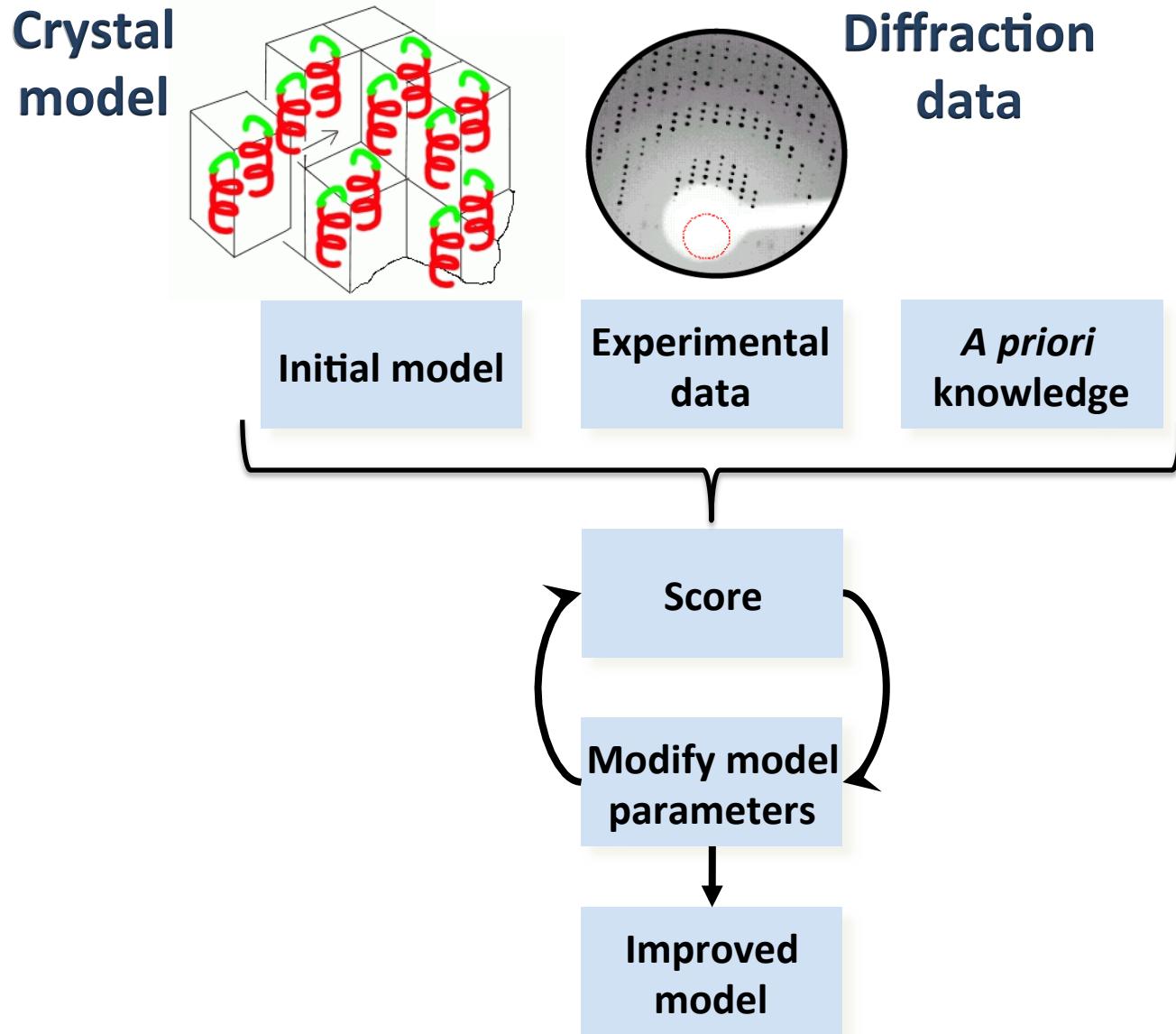
[www.phenix-online.org](http://www.phenix-online.org)

# What is Refinement?



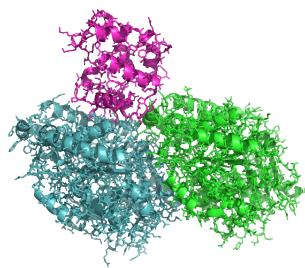
**Refinement – optimization process of fitting model parameters to experimental data**

# Refinement: Crystallography



# Refinement: Cryo-Electron Microscopy (Cryo-EM)

Atomic  
model



3D  
reconstruction  
volumes

Initial model

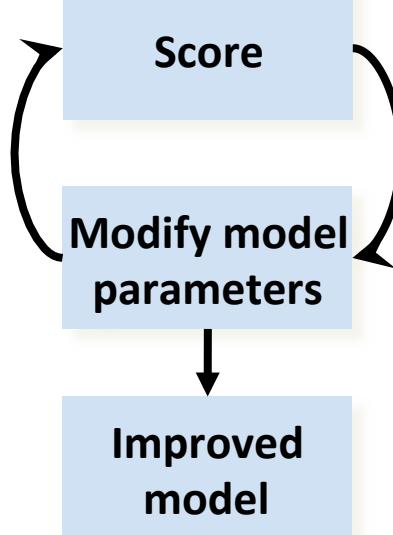
Experimental  
data

*A priori*  
knowledge

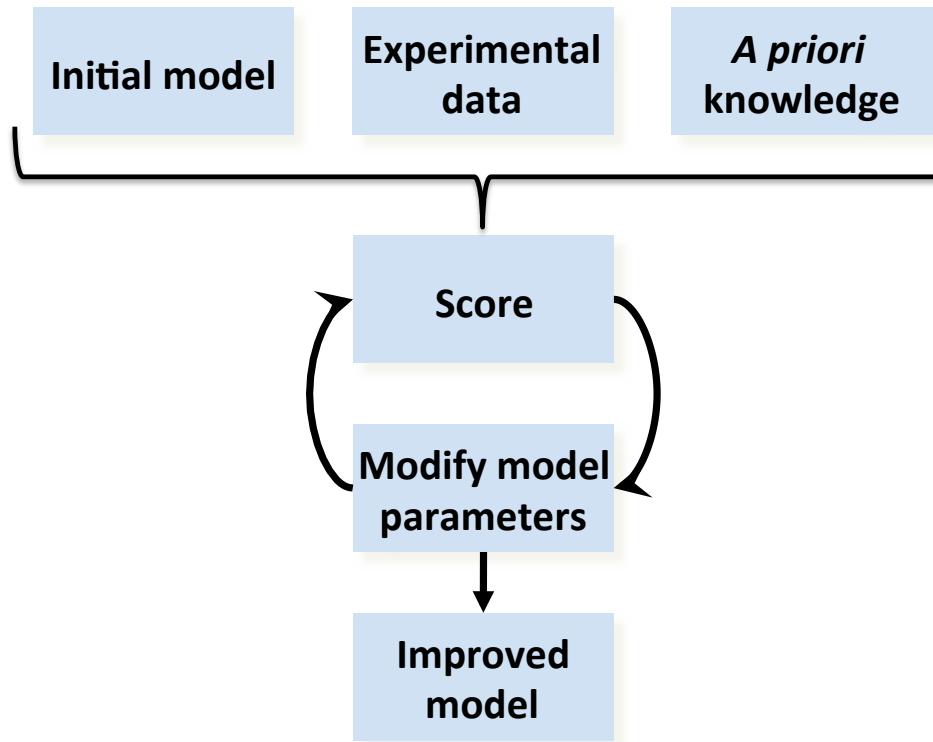
Score

Modify model  
parameters

Improved  
model



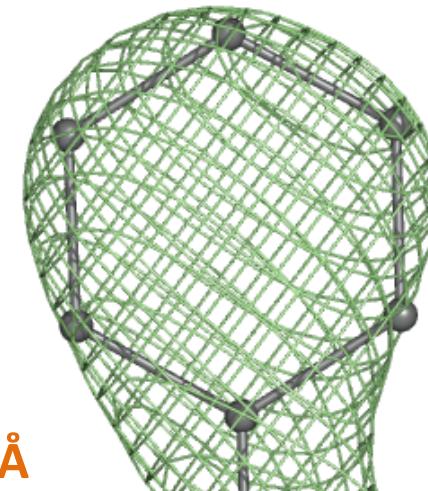
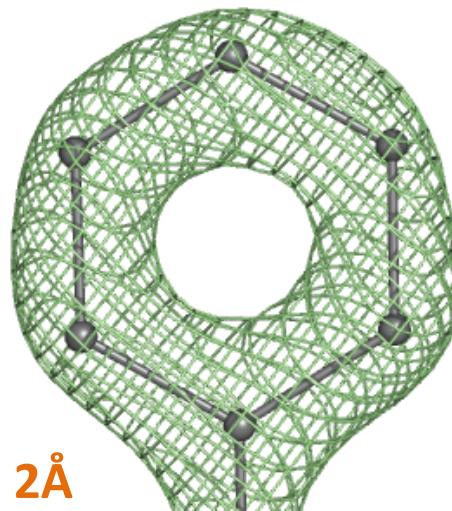
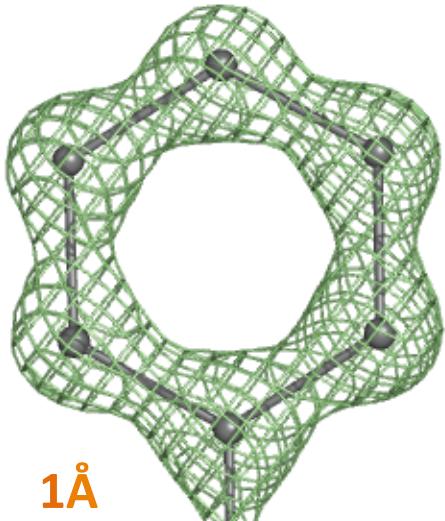
# Refinement: Score and A Priori Knowledge



**Crystallography:**  $T = T_{\text{DATA}}(F_{\text{OBS}}, F_{\text{MODEL}}) + wT_{\text{RESTRAINTS}}$

**Cryo-EM:**  $T = T_{\text{DATA}}(\rho_{\text{OBS}}, \rho_{\text{MODEL}}) + wT_{\text{RESTRAINTS}}$

# Refinement: A Priori Knowledge (Restraints)



- Lower the resolution, less detailed the map
  - Need extra information to keep correct geometry during refinement

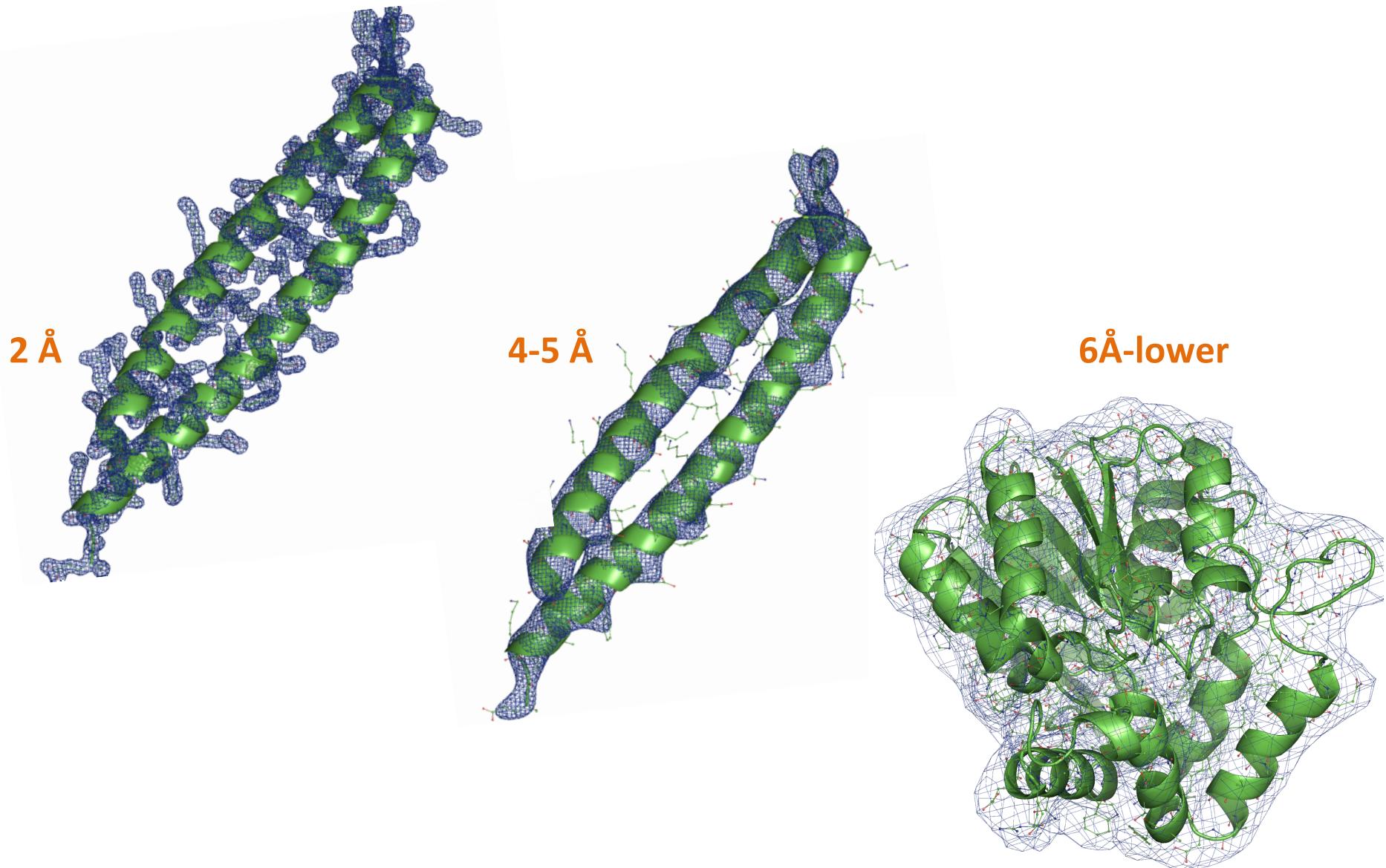
$$T_{\text{RESTRAINTS}} = T_{\text{BOND}} + T_{\text{ANGLE}} + T_{\text{DIHEDRAL}} + T_{\text{PLANARITY}} + T_{\text{NONBONDED}} + T_{\text{CHIRALITY}}$$

$$T_{\text{BOND}} = \sum_{\text{all bonded pairs}} w(d_{\text{ideal}} - d_{\text{model}})^2$$

These are standard restraints

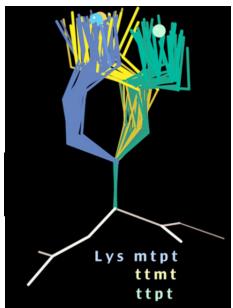
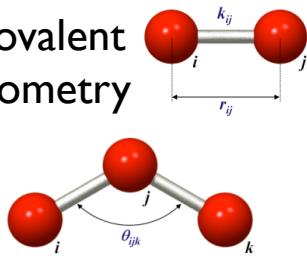
# Refinement: A Priori Knowledge (Restraints)

- Low resolution map is not sufficient to maintain secondary
  - Standard restraints are not sufficient

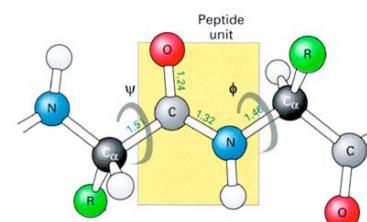
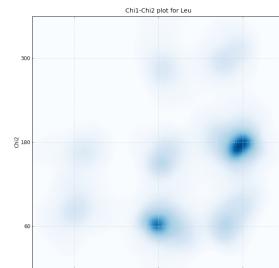


# Refinement: A Priori Knowledge (More Restraints)

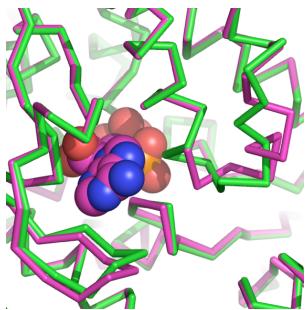
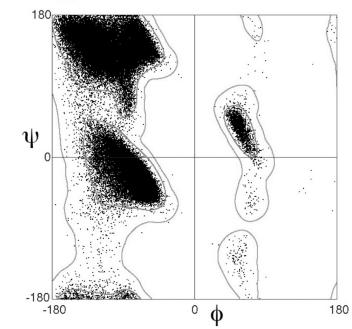
Covalent geometry



Sidechain distributions



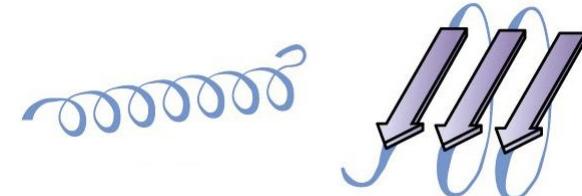
Mainchain distributions



Related structures

Internal symmetry

$$T_{\text{RESTRAINTS}} = T_{\text{BOND}} + T_{\text{ANGLE}} + \dots + T_{\text{NCS}} + T_{\text{RAMACHANDRAN}} + T_{\text{REFERENCE}} + \dots$$



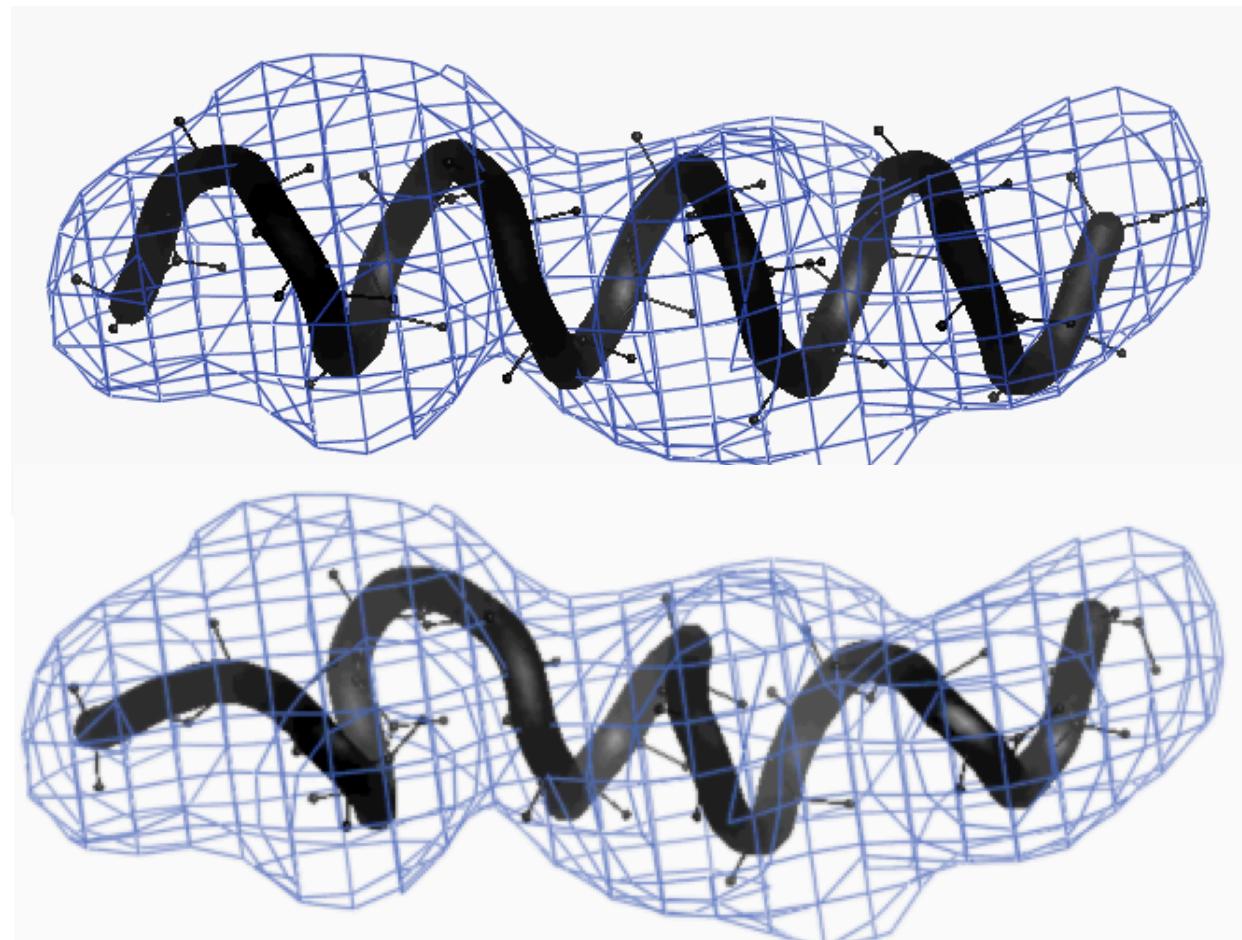
Secondary structure

## Refinement: Importance of extra restraints for low-res

- Example: refinement of a perfect  $\alpha$ -helix into low-res map
  - Using standard restraints on covalent geometry is insufficient
    - Model geometry deteriorates as result of refinement

Good helix in low-resolution map  
(before refinement)

Distorted helix in low-resolution map  
(after refinement,  
still fits the map!)

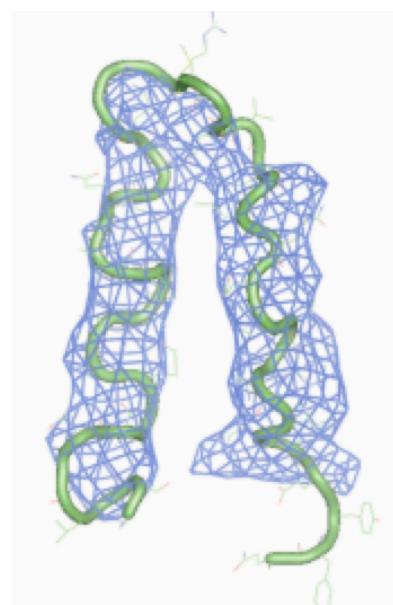


# Refinement: Importance of extra restraints for low-res

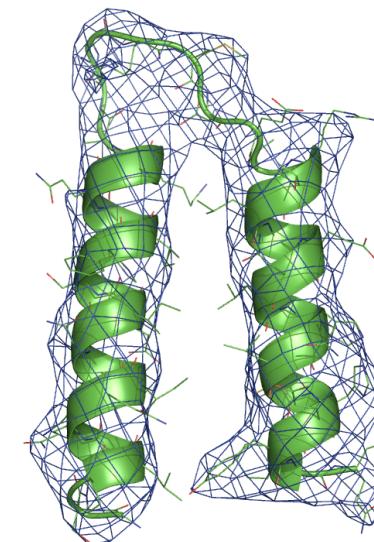
- Example: refinement of distorted model into low-res map
  - Using standard restraints on covalent geometry is insufficient
    - Good model-to-map fit, poor geometry



Before refinement

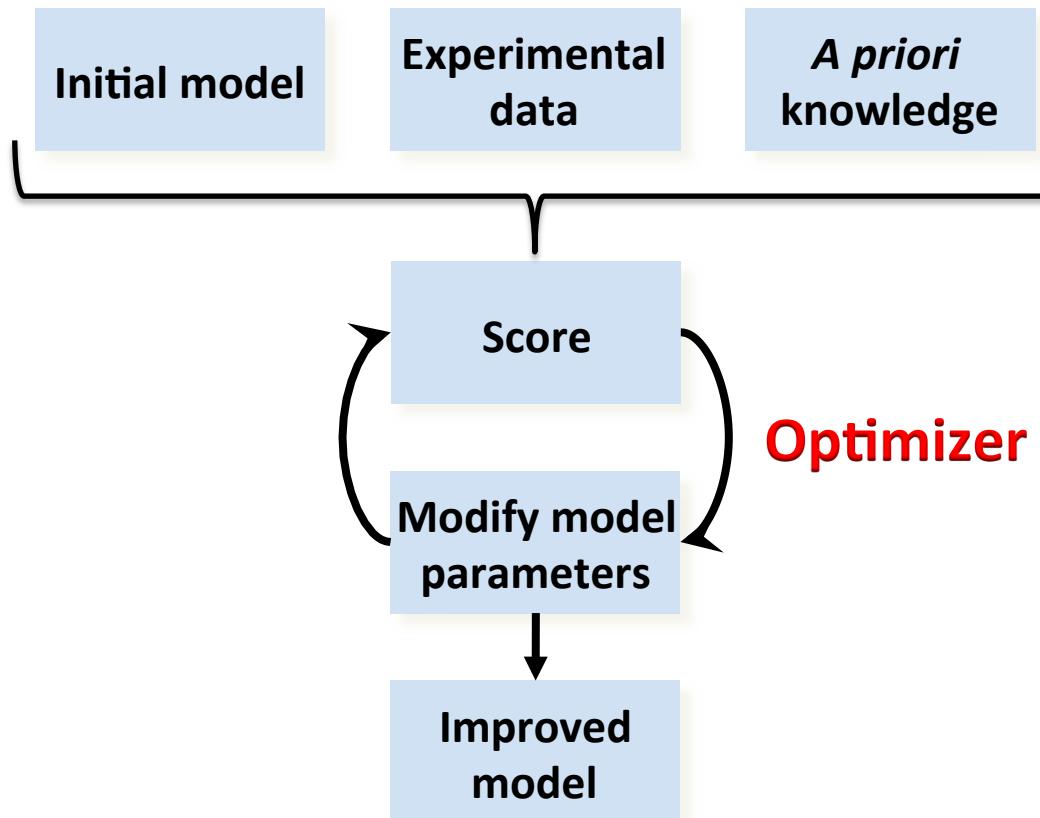


After refinement

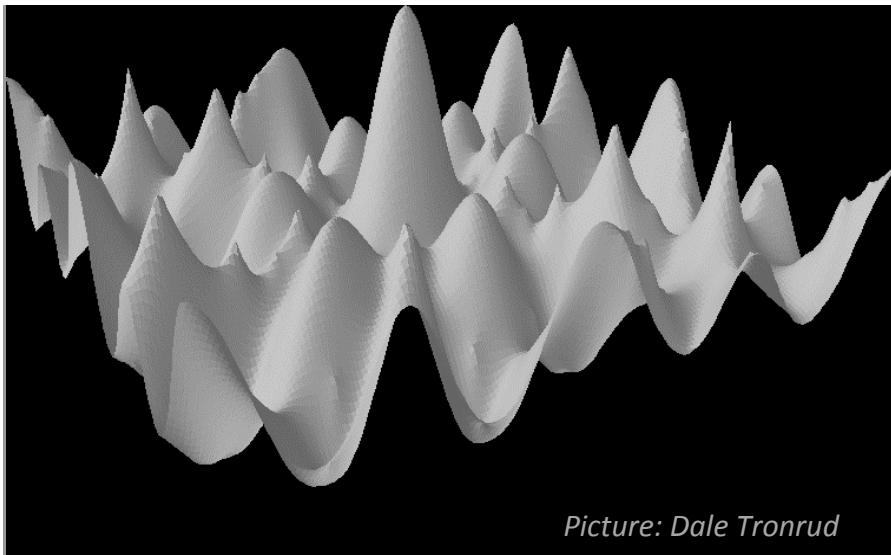


True (expected)  
result

# Refinement: Optimizer



# Refinement: Optimizer

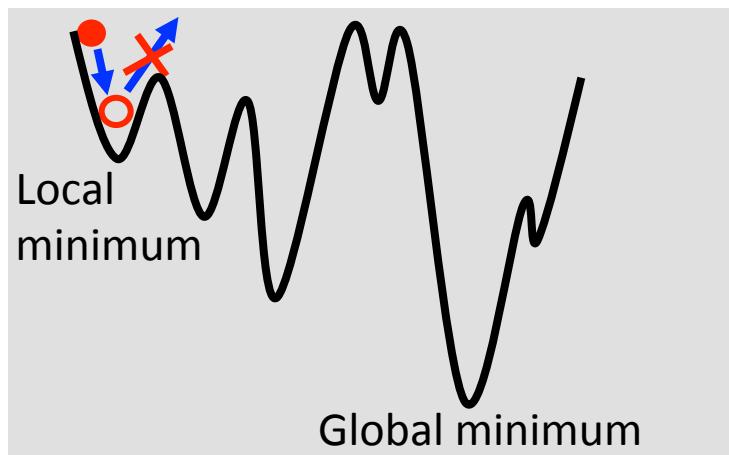


Global  
minimum

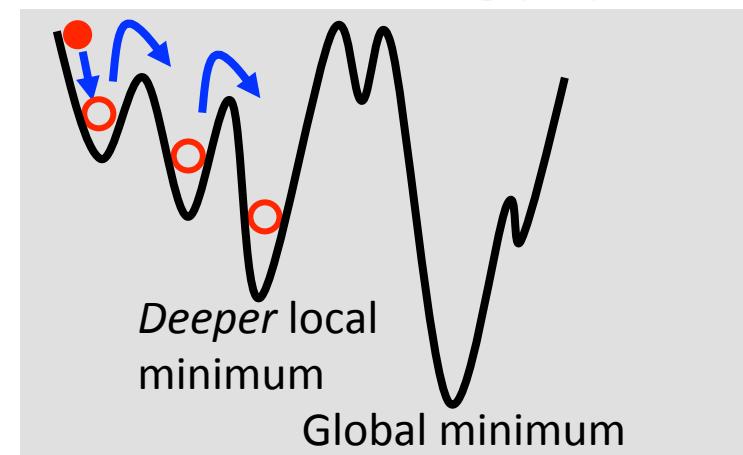
- Refinement goal is to reach the global minimum
  - Rarely achieved -> limited convergence radius
    - Refinement result strongly depends on starting point
    - Various optimization methods are used

# Refinement: Optimizer

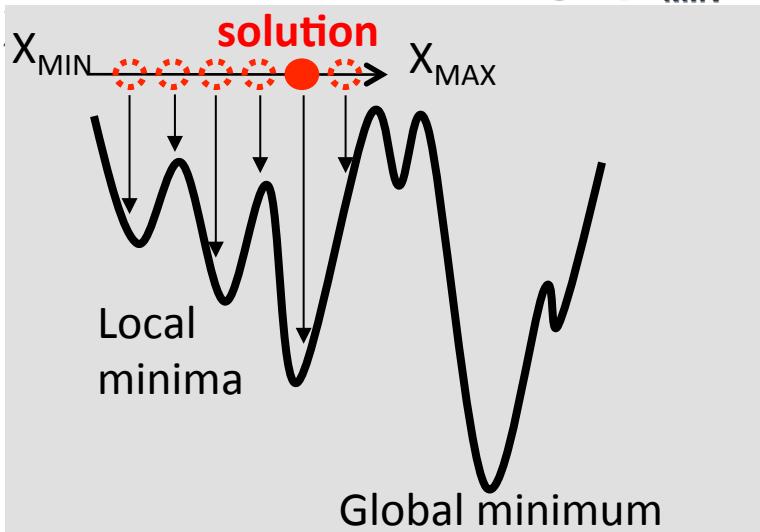
- Gradient-driven minimization



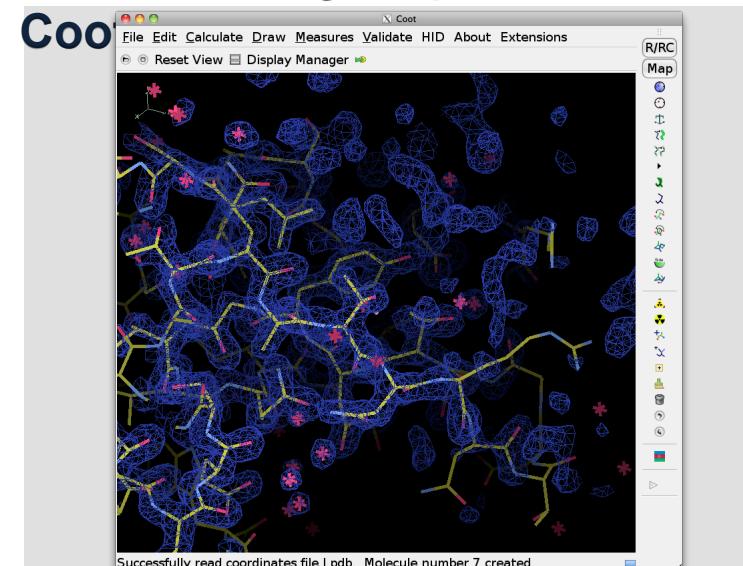
- Simulated annealing (SA)



- Grid search (Sample parameter space within known range [ $X_{MIN}$ ,  $X_{MAX}$ ])

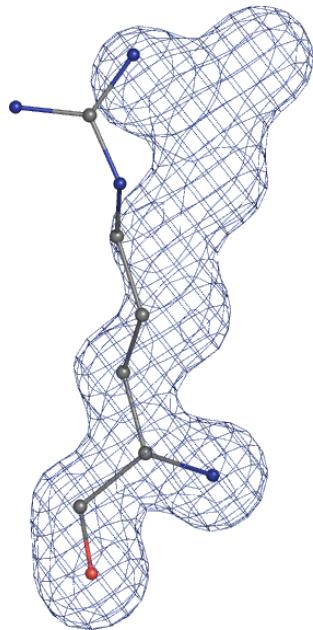


- Hands & eyes (via Coot)

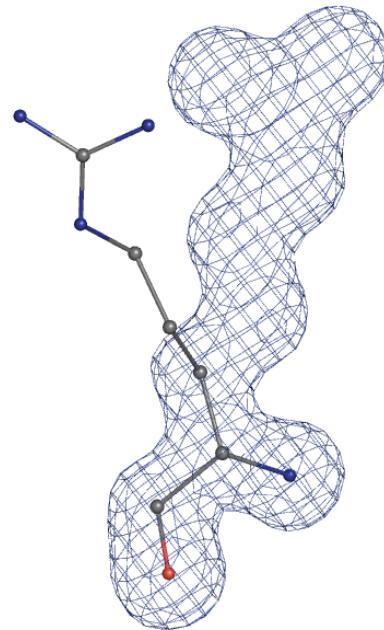


# Refinement: Optimizer

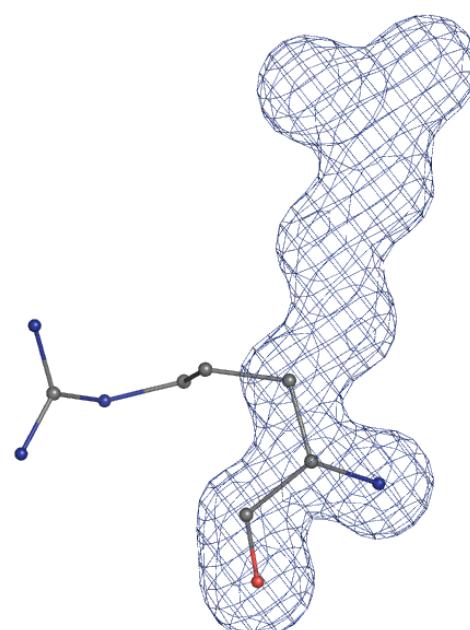
Minimization



Simulated Annealing



Real-space grid search



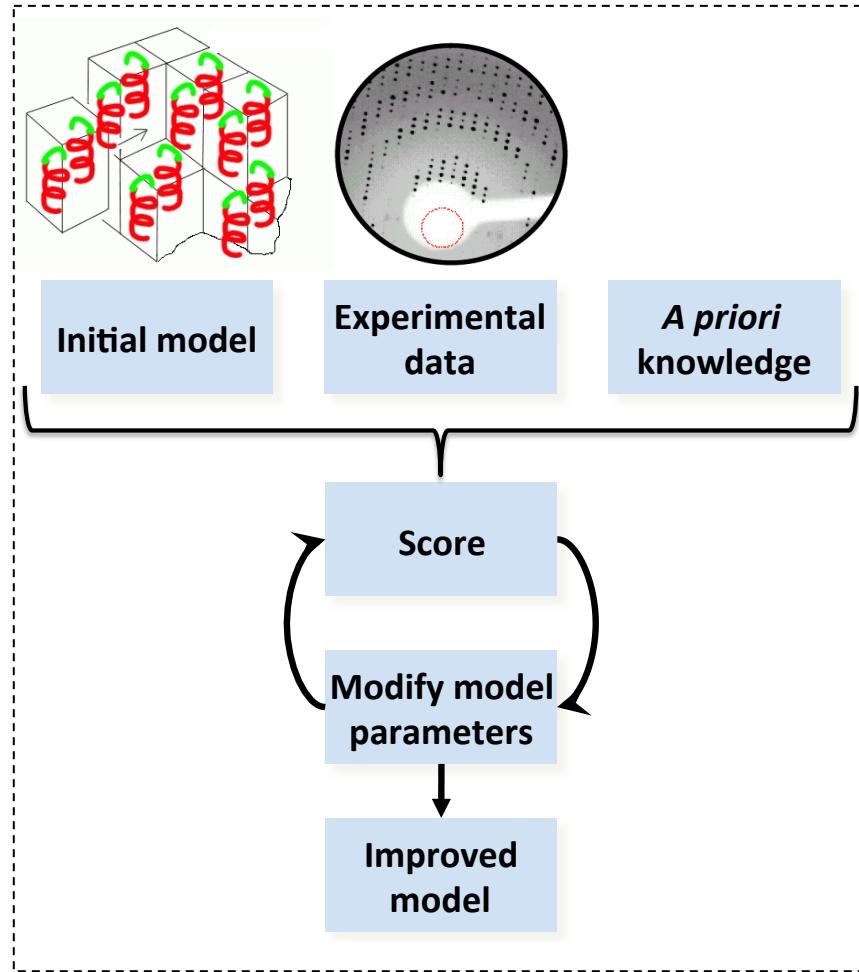
Minimization or SA can fix it

Beyond convergence radius of minimization

Beyond convergence radius of minimization and SA

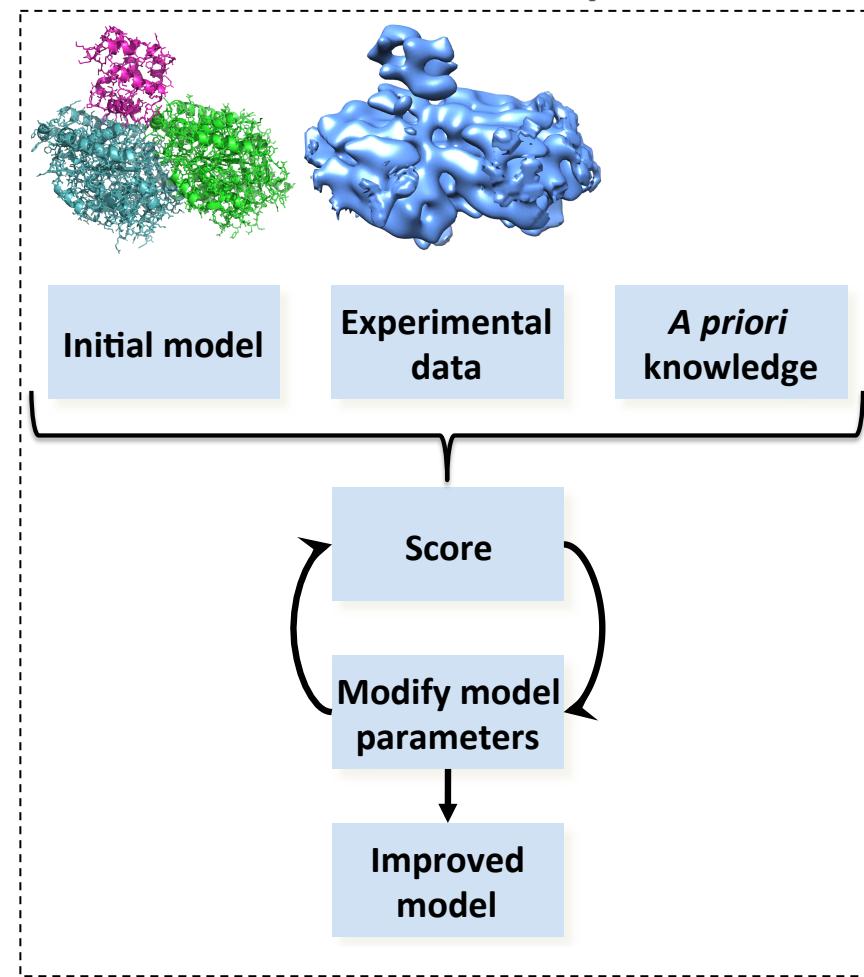
# Refinement tools in Phenix

## Crystallography (or reciprocal space refinement)



**phenix.refine**

## Cryo-EM (or real space refinement)



**phenix.real\_space\_refine**

# Refinement tools in Phenix

PHENIX home

Citations Reload last job

Coot PyMOL KiNG Other tools Ask for help

**Actions Job history**

**Projects**

Show group: All groups Manage...

Select Delete New project Settings

ID	Last modified	# of jobs	R-free
ringer	Sep 07 2016 05:37 ...	2	---
tmp2	Sep 07 2016 05:23 ...	1	---
5gnn	Sep 07 2016 08:42 ...	1	---
debug1	Sep 05 2016 10:51 ...	2	0.0086
tmp4	Aug 18 2016 07:23 ...	2	---
testing	Aug 11 2016 01:54 ...	1	---
mich	Jul 29 2016 12:47 ...	1	---
almu	Jul 28 2016 10:58 ...	1	---
rchen	Jul 22 2016 11:10 ...	1	---
milya	Jul 15 2016 12:36 ...	2	---
SEM	Jul 14 2016 05:20 ...	1	0.1570

**Data analysis**

**Experimental phasing**

**Molecular replacement**

**Model building**

**Refinement**

**phenix.refine**  
Automated X-ray and/or neutron refinement

**Real-space refinement**  
Automated real-space refinement

**Neutron refinement [alpha]**  
Alternate phenix.refine interface customized for neutron refinement

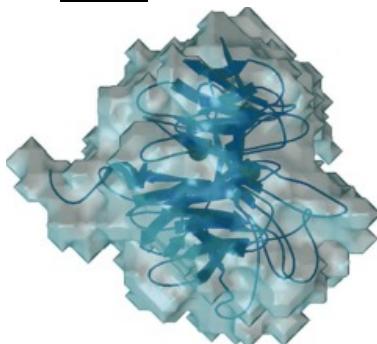
**DEN refinement [alpha]**  
Deformable elastic network refinement using sparse resolution and molecular replacement strategy

Current directory: /Users/pafonine/Desktop/work/tmp Browse...

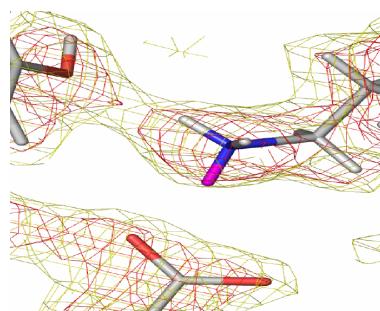
PHENIX version dev-svn-000 Project: 5gnn

# Phenix.refine (available in Phenix since 2005)

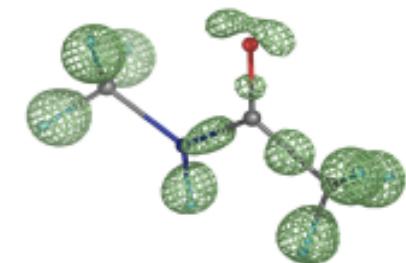
## Low



## Medium and High



## Subatomic



- Restrained/constrained refinement of individual parameters
- Automatic water update

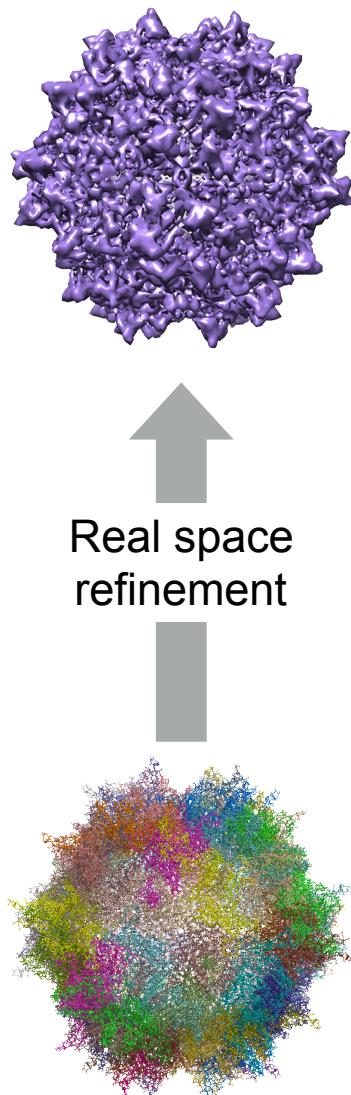
- Bond density model
- Unrestrained refinement
- FFT or direct
- Explicit hydrogens

- Group ADP refinement
- Rigid body refinement
- Torsion Angle dynamics
- Reference model
- Ramachandran plot restraints
- Secondary structure restraints

- Automatic NCS restraints, constraints
- Simulated Annealing
- Automatic side chain rotamer fixing
- Occupancies (individual, group, automatic constraints for alternative conformations)
- Various targets: LS, ML, MLHL,...
- Dual (real/reciprocal) space refinement

- TLS refinement with automated TLS groups identification
- Use hydrogens at any resolution
- Refinement with twinned data
- X-ray, Neutron, joint X-ray + Neutron

# Phenix.real\_space\_refine (available in Phenix since 2013)



VOLUME FOUR JULY MMXIII

**COMPUTATIONAL CRYSTALLOGRAPHY NEWSLETTER**

**ENSEMBLE REFINEMENT, CABLAM**

**Table of Contents**

• PHENIX News	28
• Crystallographic meetings	29
• Expert Advice	
• Fitting tips #6 - Potential misfitting by a switch of sidechain vs mainchain	30
• FAQ	32
• Short Communications	
• <i>CaBLAM</i> : Identification and scoring of disguised secondary structure at low resolution	33
• Structural Classification of Allergen IgE Epitopes by Hierarchical Clustering	36
• New Tool: <i>phenix.real_space_refine</i>	43
• <i>cctbx</i> tools for transparent parallel job execution in Python. III. Remote access	45
• <i>phenix.ensemble_refinement</i> : a test study of apo and holo BACE1	51

**Editor**  
Nigel W. Moriarty, [NWMoriarty@LBL.Gov](mailto:NWMoriarty@LBL.Gov)

**PHENIX News**

**New programs**

FEM: Feature Enhanced Maps (Pavel V. Afonine)

Interpretation of a crystallographic map is a means of obtaining an atomic representation of a crystal structure or the map itself may serve as the crystal model. There are number of factors that affect quality of crystallographic maps that in turn affect difficulty (or even feasibility) of their interpretation and quality of resulting model of crystal structure, and include:

- finite resolution of measured reflections;
- incompleteness of data (missing reflections within the resolution range of the measured data);
- experimental errors in measured reflections;
- errors in atomic model parameters.

These factors a) result in artificial peaks in the map that may be confused with the signal and therefore erroneously interpreted in terms of atomic model, b) introduce noise that may obscure the signal and c) may distort the signal in various ways.

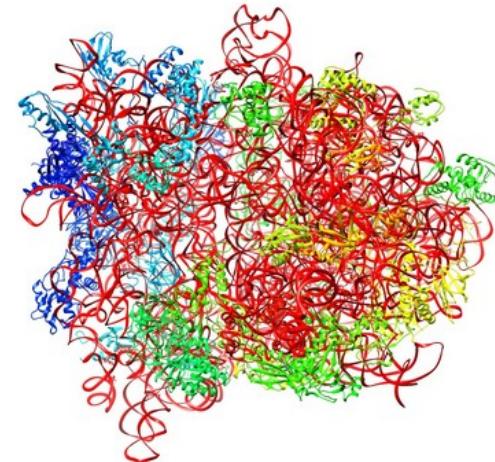
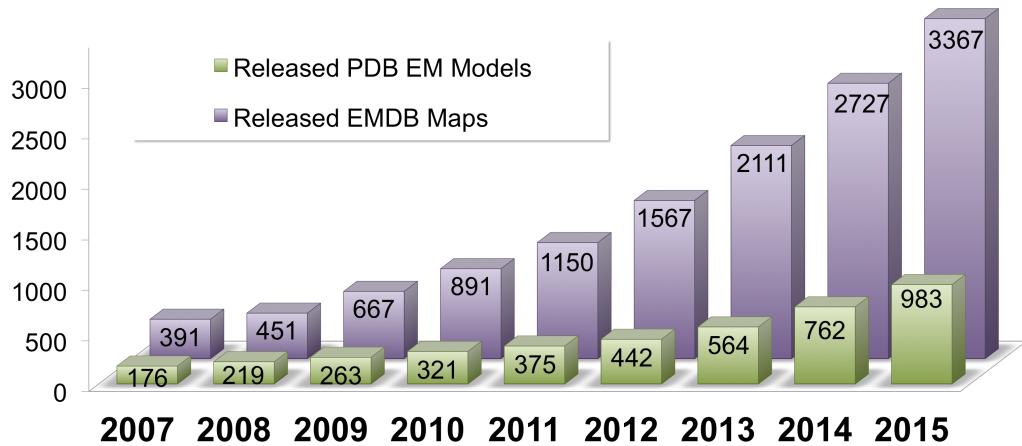
Another fundamentally different contributor to the difficulty of map interpretation is that not all the signal has the same strength. For example, a strong signal arising from a heavy atom derivative may easily obscure a very weak signal (that may be at or below the noise level) arising from a partially occupied very mobile ligand or residue side chain alternative conformation or even hydrogen atoms.

The Computational Crystallography Newsletter (CCN) is a regularly distributed electronically via email and the PHENIX website, [www.phenix-online.org/newsletter](http://www.phenix-online.org/newsletter). Feature articles, meeting announcements and reports, information on research or other items of interest to computational crystallographers or crystallographic software users can be submitted to the editor at any time for consideration. Submission of text by email or word-processing files using the CCN templates is requested. The CCN is not a formal publication and the authors retain full copyright on their contributions. The articles reproduced here may be freely downloaded for personal use, but to reference, copy or quote from it, such permission must be sought directly from the authors and agreed with them personally.

Computational Crystallography Newsletter (2013). Volume 4, Part 2.

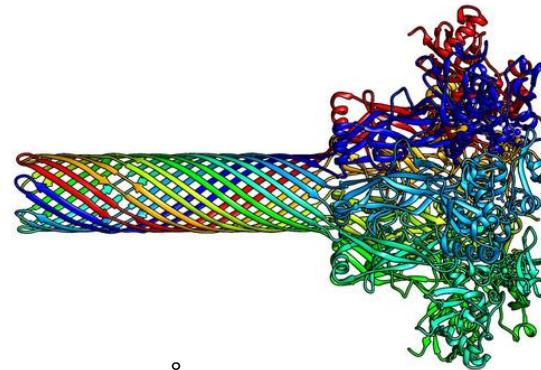
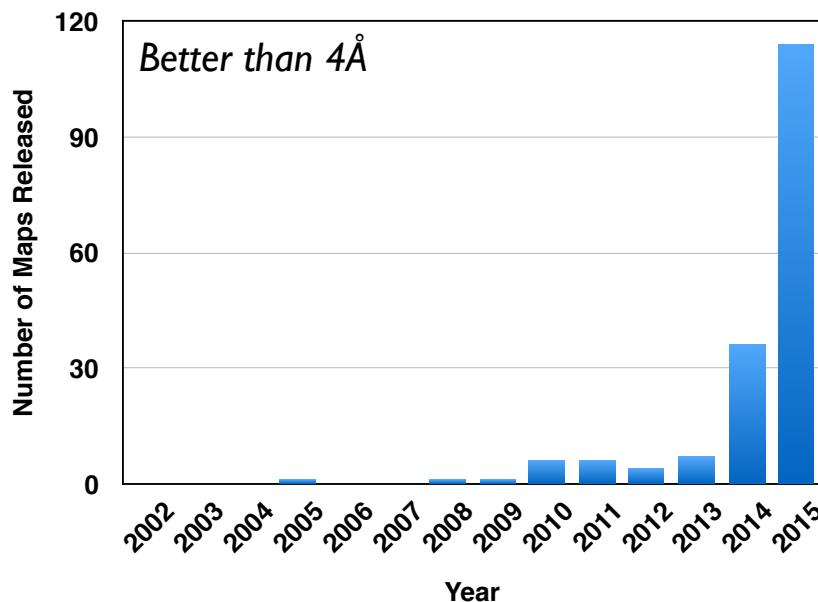
28

# Rapid Growth of Higher Resolution Cryo-EM



2.9 Å Ribosome (5AFI)

Fischer et al., *Nature* 520: 567 (2015)



2.9 Å ATPAP (3J9C)

Jiang et al., *Nature* 521: 545 (2015)

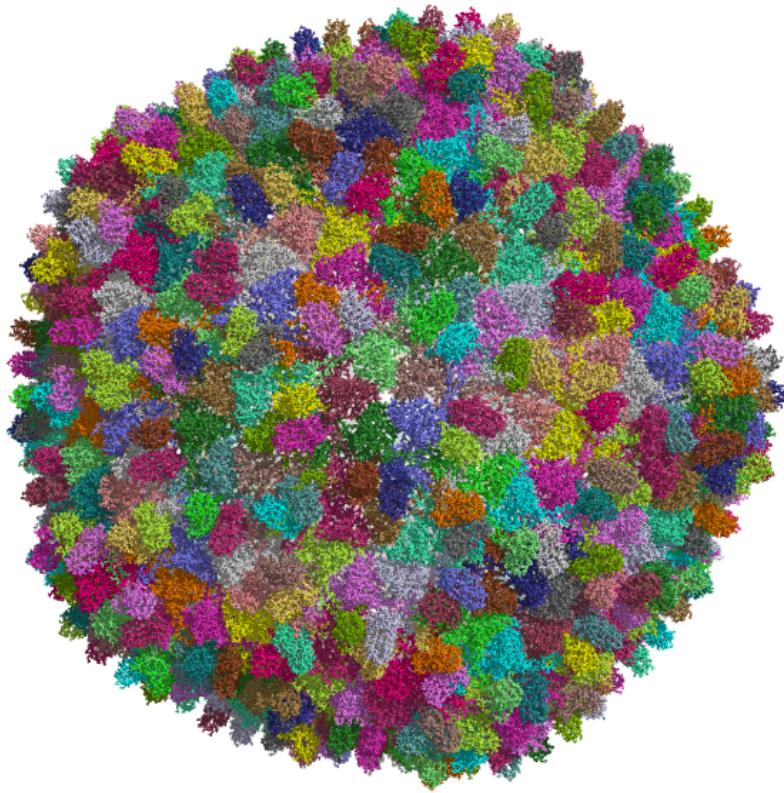
# Challenges

- **Automated model building**
  - What is magnification of the map? (can be 10% uncertainty)
  - What is optimal sharpening of the map?
  - What is the region containing the molecule?
  - Low and variable resolution across maps
- **Structure optimization (refinement)**
  - Low and variable resolution across maps
  - Large molecules
  - Poor initial models
- **Validation**
  - How to validate a model against low resolution maps
    - Validation metrics and tools

# More Challenges

- People
  - Many hard-core crystallographers doing cryo-EM now
    - Stereotypes
      - Rfree
      - Thinking of cryo-EM maps in terms of structure factors
- Handling data
  - Cryo-EM:
    - Maps can be “massaged” many times as part structure solution (cut-out, averaged, masked, sharpened, blurred...)
      - B factors?
  - Crystallography:
    - Fobs or lobs are never touched!
  - Storing models:
    - Too large models may not fit PDB format
    - CRYST1 records?

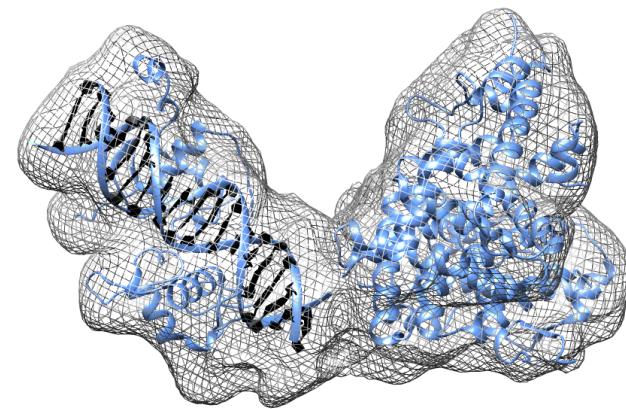
# Model Refinement - Challenges



Resolution 4.5 Å

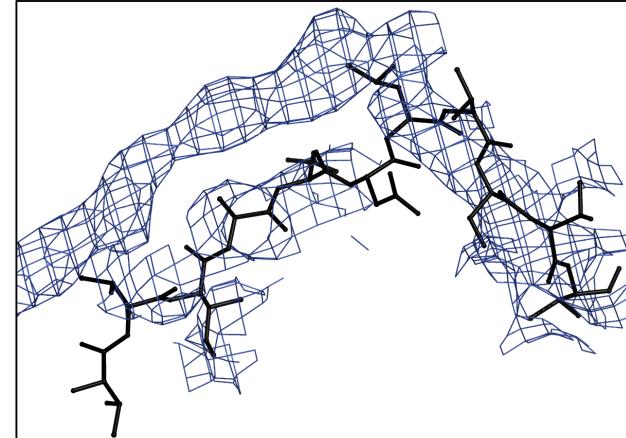
840 chains, 187,320 residues 1,443,960 atoms

Size



Resolution: 11.6 Å

Wide Resolution Range



User data, resolution: 3.8 Å

Poor Initial Fit

## PDB - Poor geometry: Refinement Tools are Needed

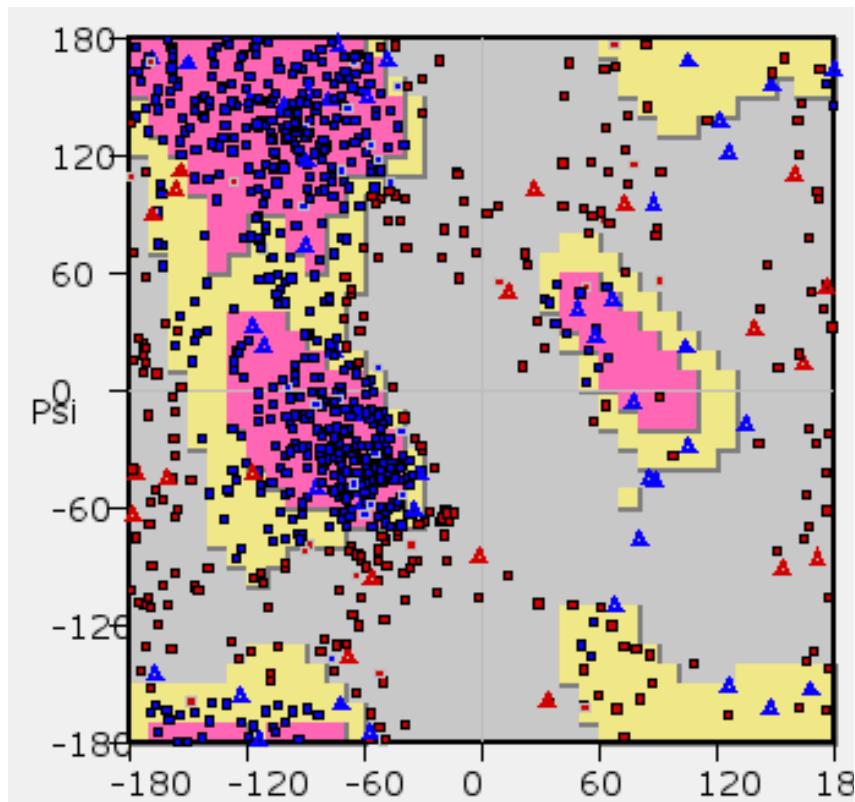
- PDB survey of all deposited cryo-EM models (~750)

% Ramachandran outliers	No. Structures
3	546
8	82
14	24
20	13
25	4

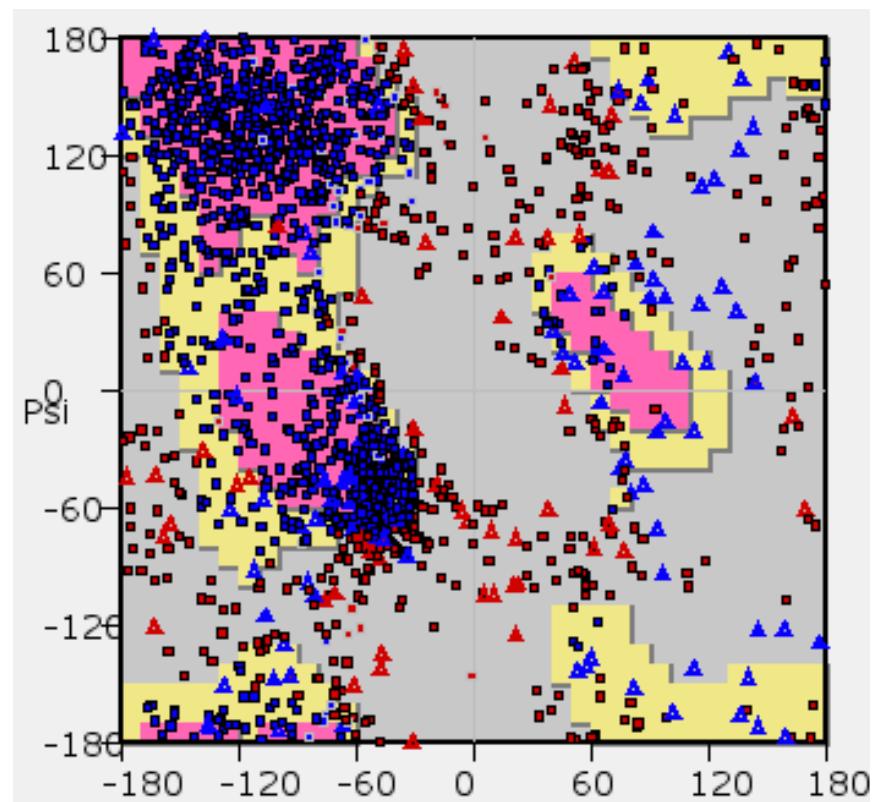
**Goal: 0%**

# PDB - Poor geometry: Refinement Tools are Needed

PDB: 3J0?, resolution 10 Å



PDB: 3J2?, resolution 3.6 Å



# PDB - Poor geometry: Refinement Tools are Needed

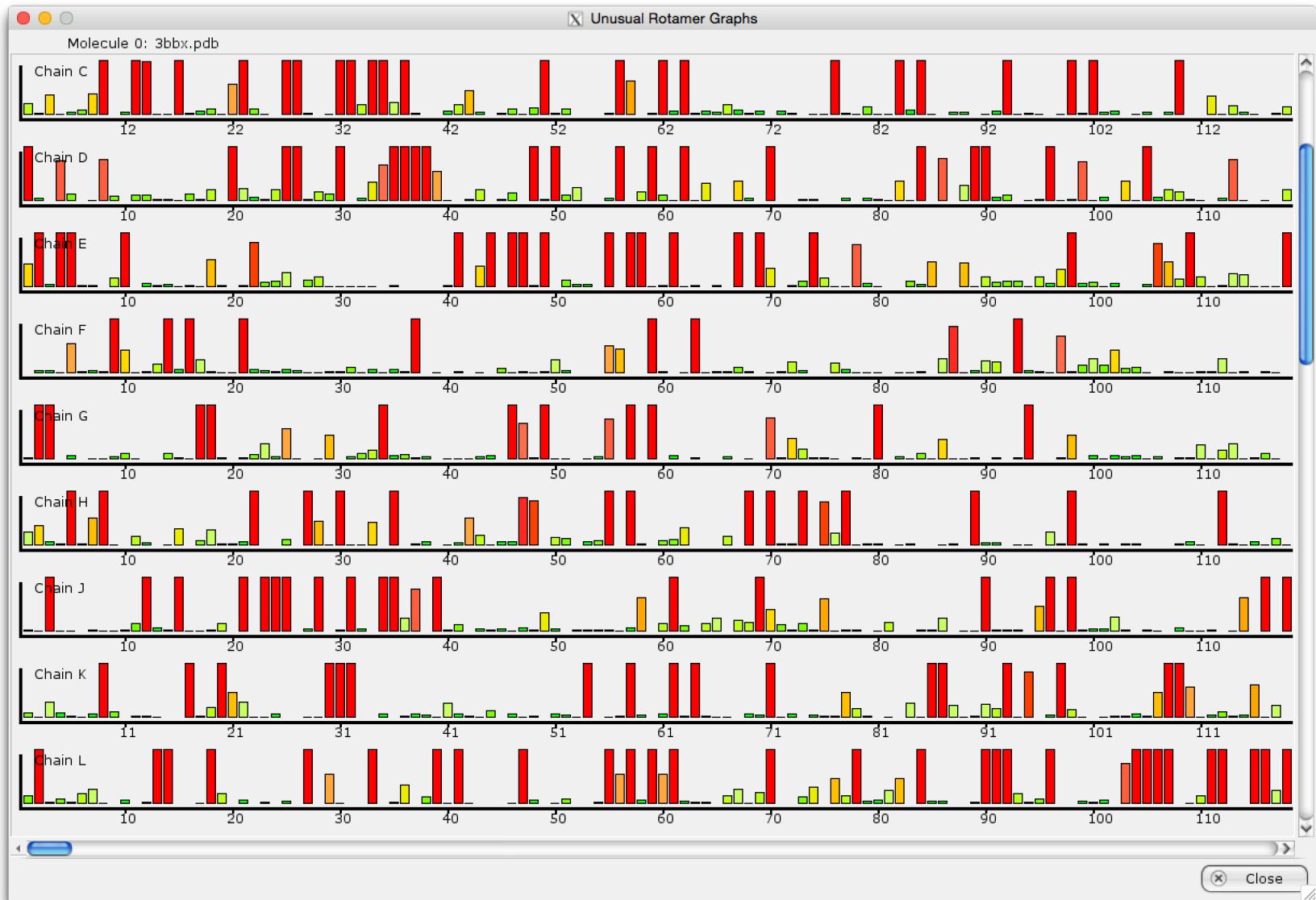
- PDB survey of all deposited cryo-EM models

% Rotamer outliers	No. Structures
5	505
15	115
24	37
34	7
44	5

**Goal: 0%**

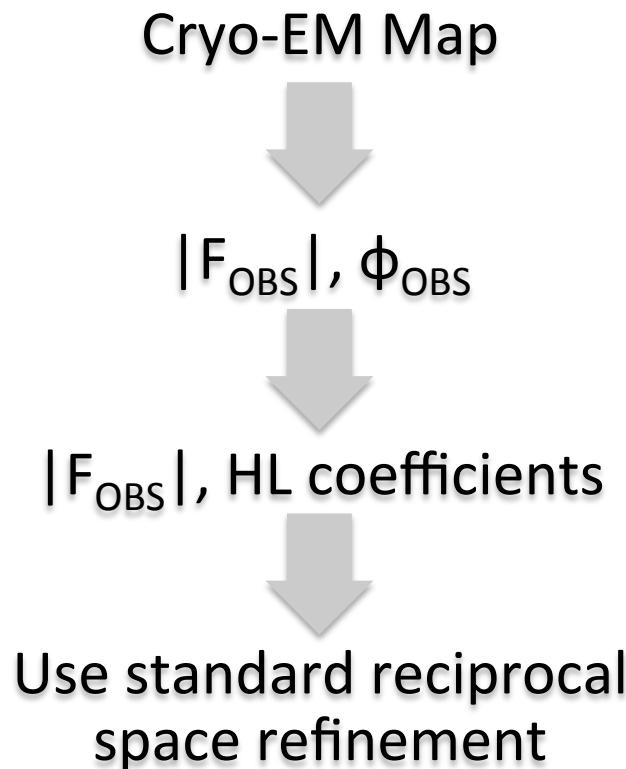
# PDB - Poor geometry: Refinement Tools are Needed

- Rotamer outliers per chain (PDB code: 3BBX)



# Refinement – software bottlenecks

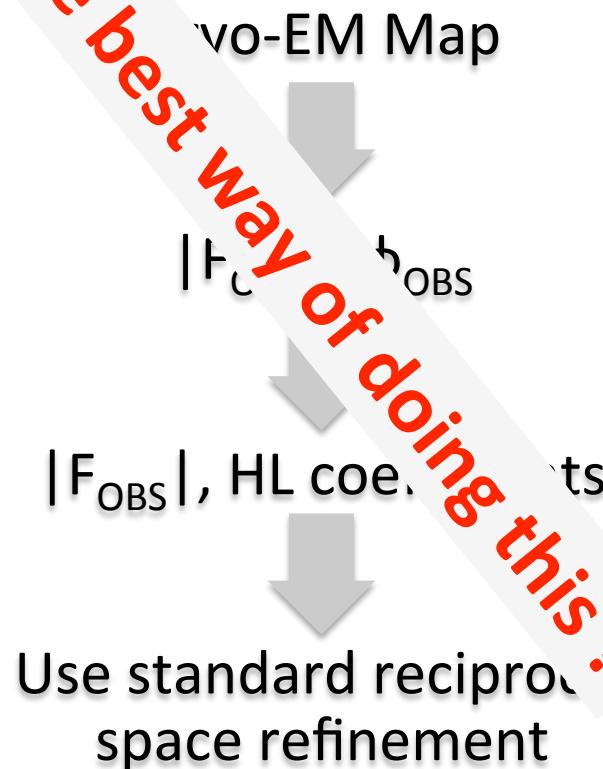
- Lack of dedicated tools for refinement of atomic models into Cryo-EM maps
  - Programs designed for crystal structure refinement are used to refine against cryo-EM maps



# Refinement – software bottlenecks

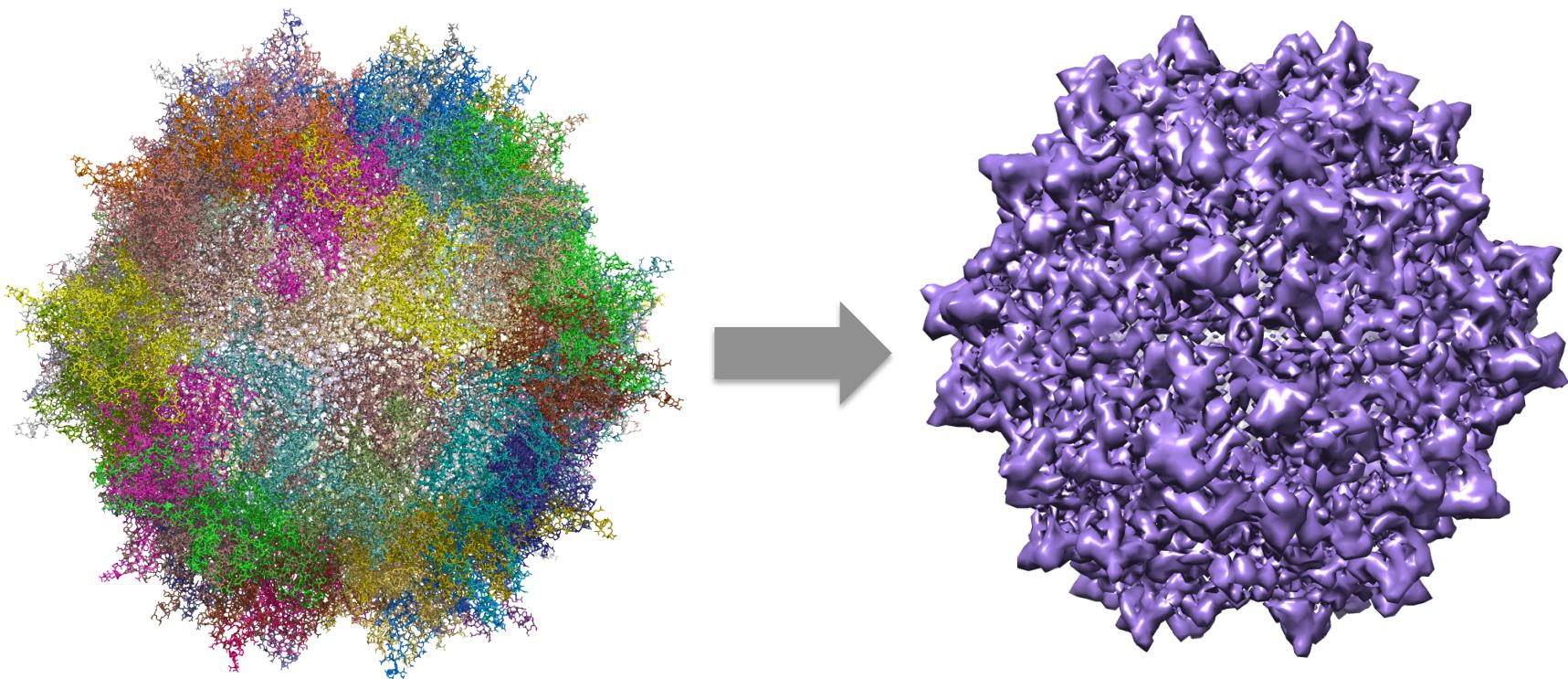
- Lack of dedicated tools for refinement of atomic models into Cryo-EM maps

- Programs designed for crystal structure refinement are used to refine against cryo-EM maps



## Phenix.real\_space\_refine

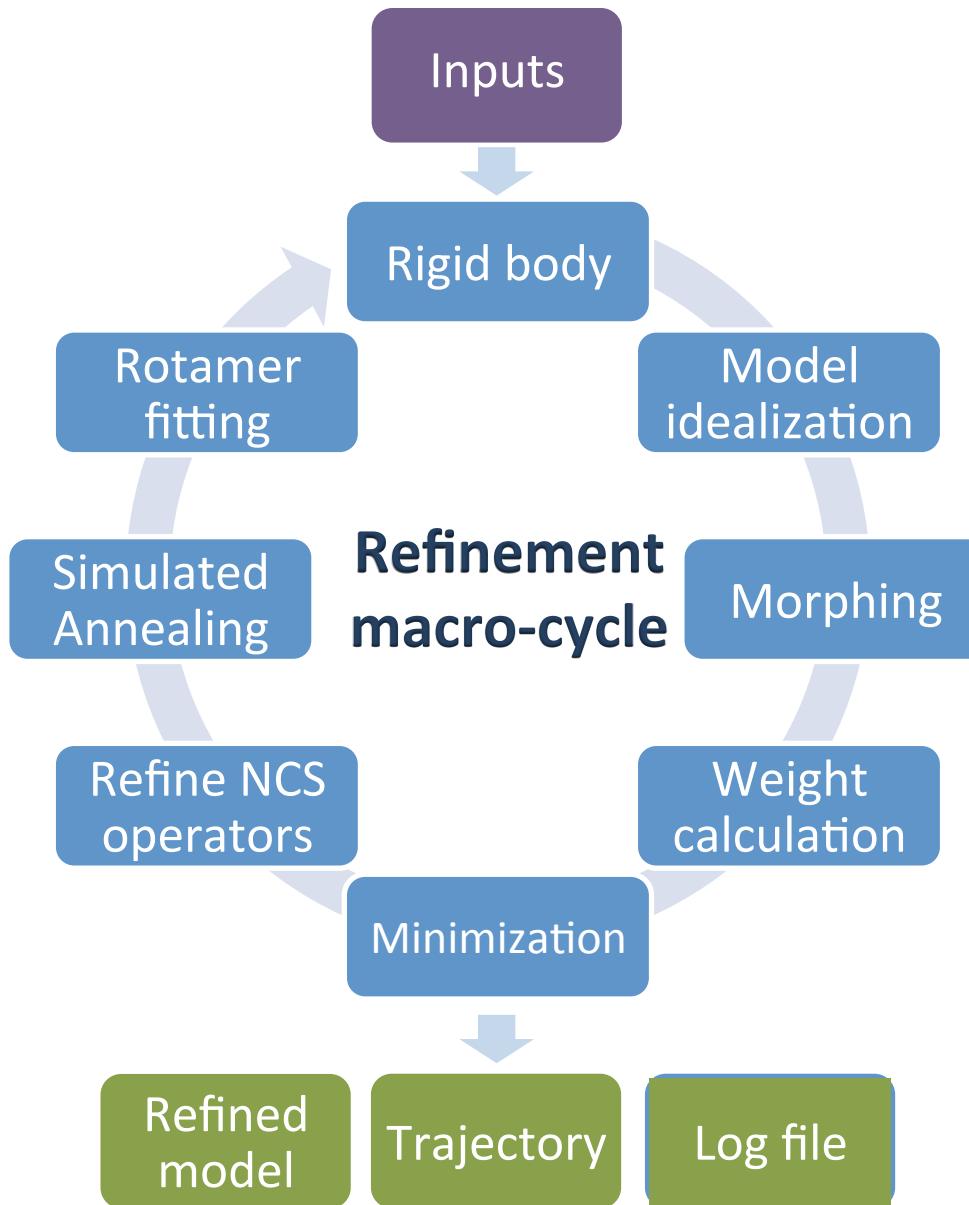
- Direct refinement against the map
  - Cryo-EM map is closest to experimental data: direct refinement against the map



## **Phenix.real\_space\_refine – start around 2013**

- Best model-map fit. Any map: X-ray, neutron, EM. Any resolution
- Refined models: no poor validation metrics
- Fast (minutes – a few hours, not days or many hours!)
- Large convergence radius
- Easy to use: map and model in, refined model out!

# Phenix.real\_space\_refine – state-of-the-art



# Phenix.real\_space\_refine – model idealization

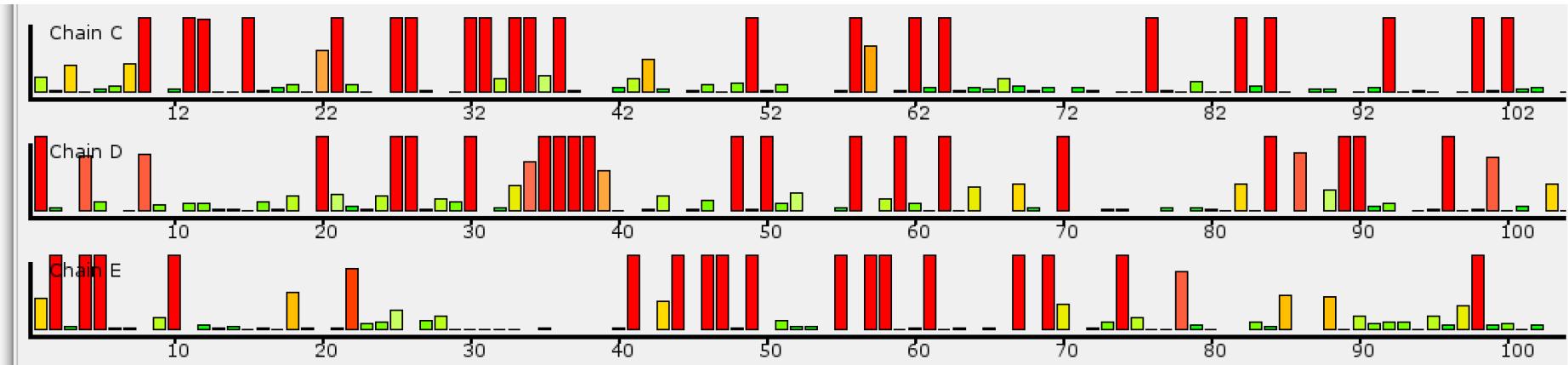
- Goal
  - Eliminate all geometry outliers
  - Move atoms as little as possible from start position
    - Idealized model within convergence of refinement
- Why?
  - Refinement may not be able to refine a model with lots of bad geometries

Oleg Sobolev

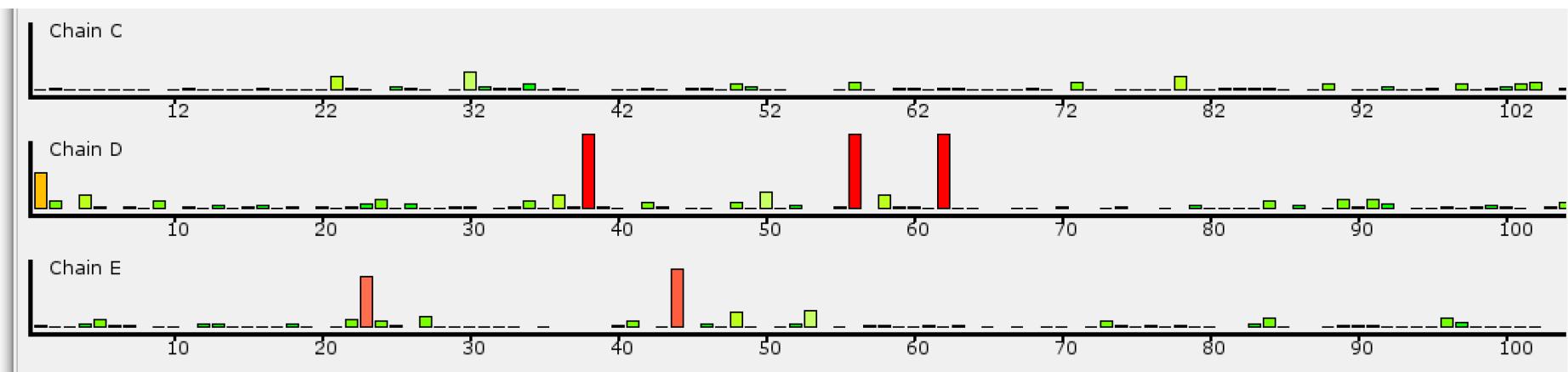
# Phenix.real\_space\_refine – model idealization

- PDB code: 3BBX

Before



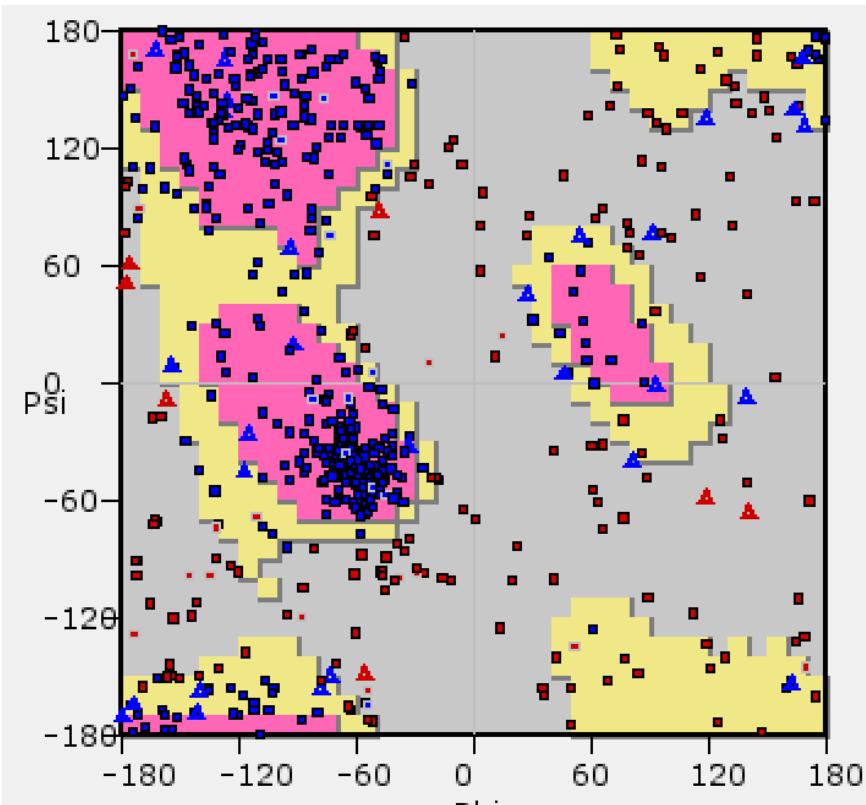
After



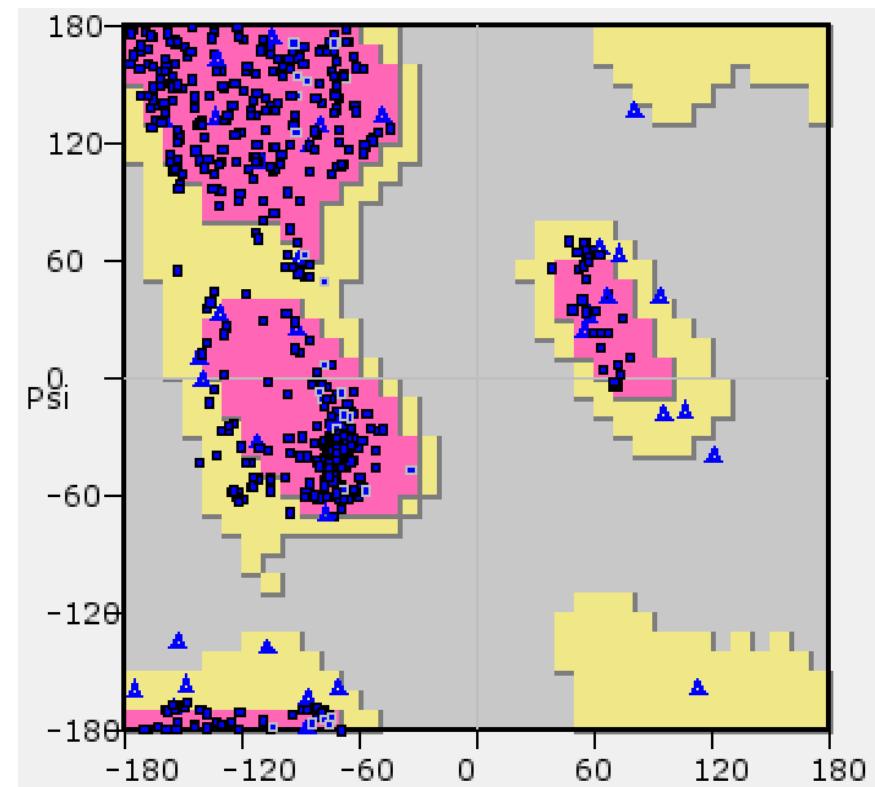
# Phenix.real\_space\_refine – model idealization

- PDB code: 3J25

Before



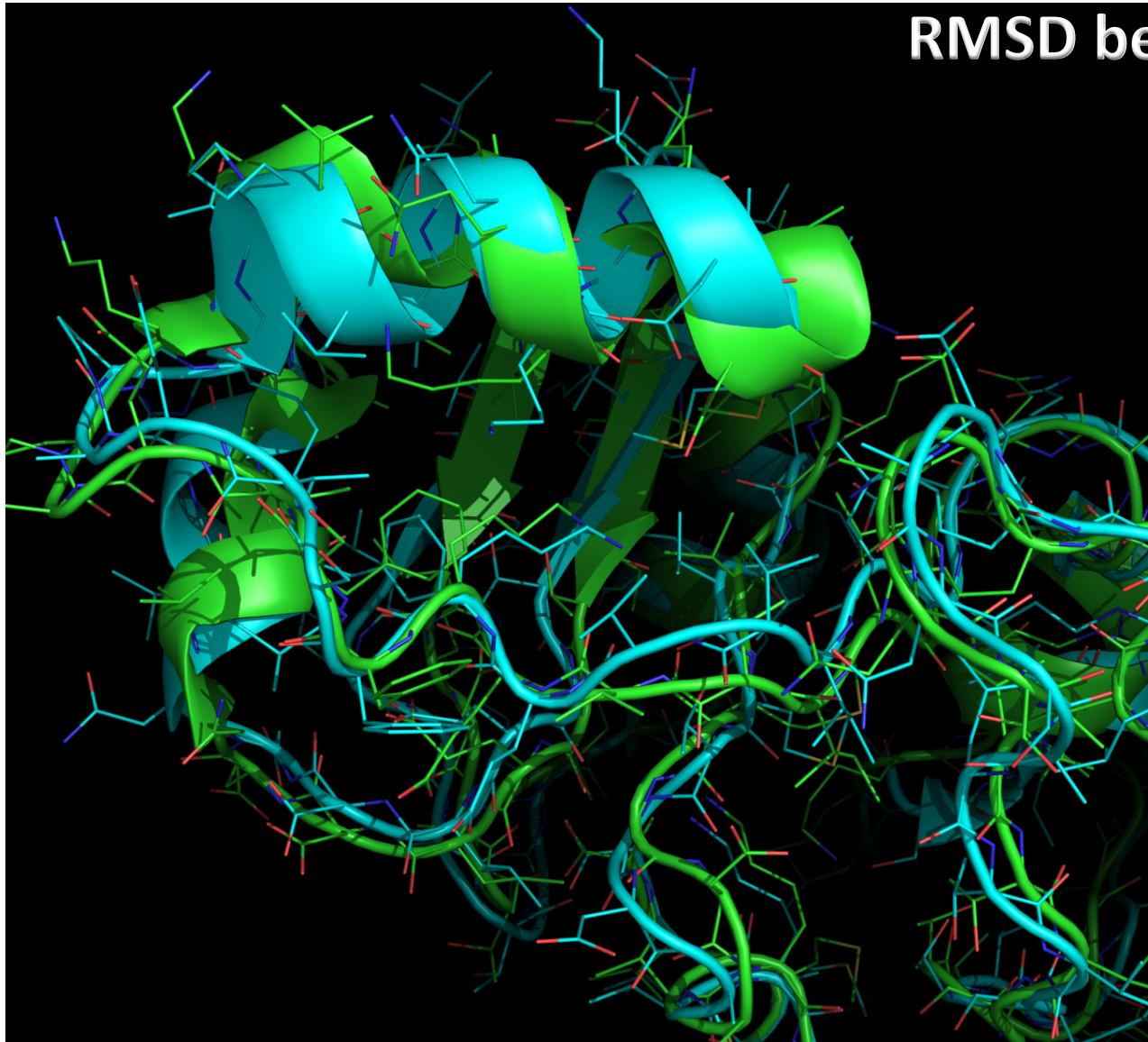
After



# Phenix.real\_space\_refine – model idealization

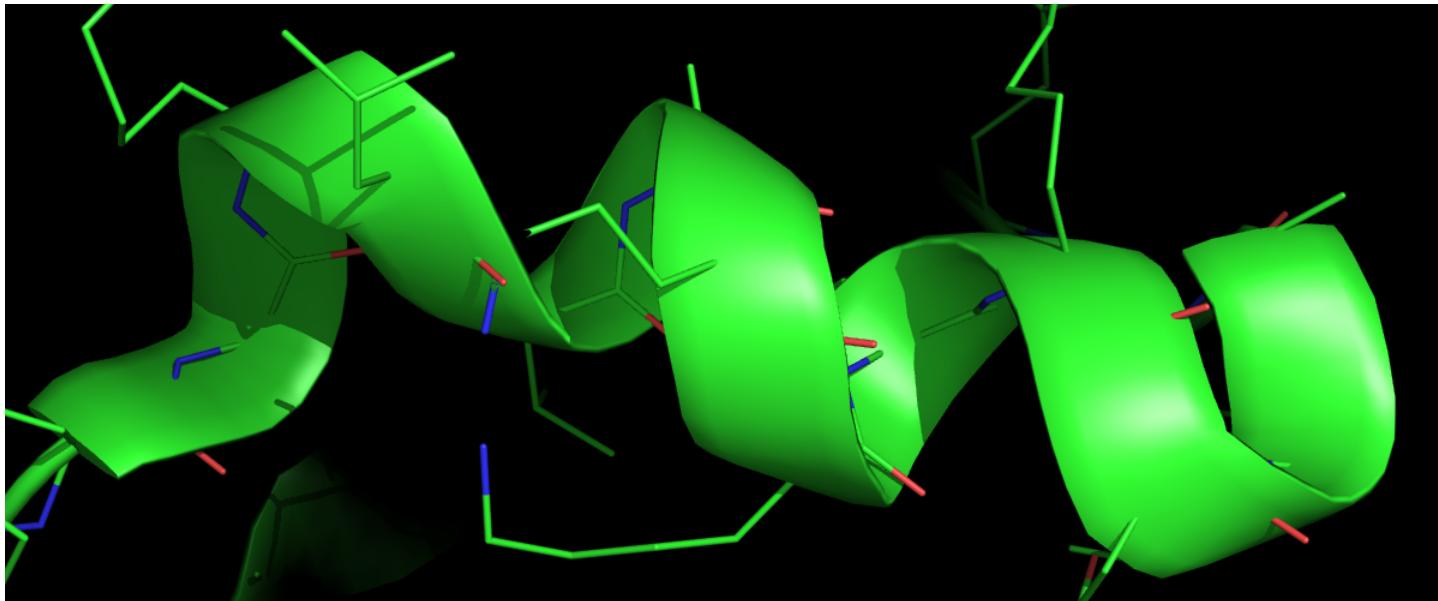
Before and after idealization

RMSD between two models  
less than 1.5Å

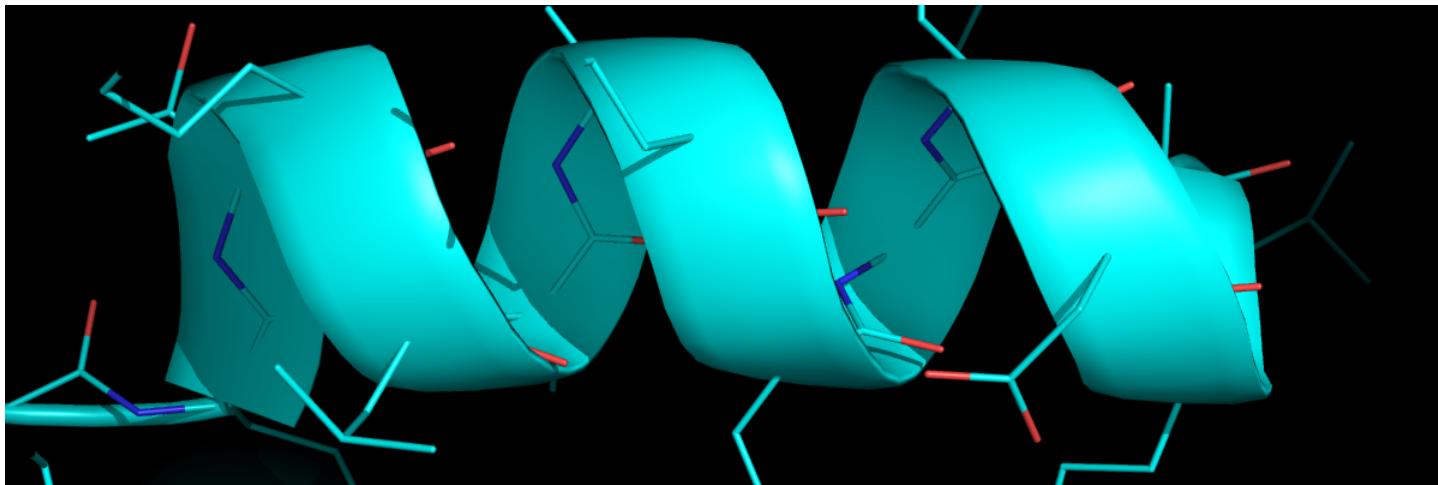


# Phenix.real\_space\_refine – model idealization

Before...

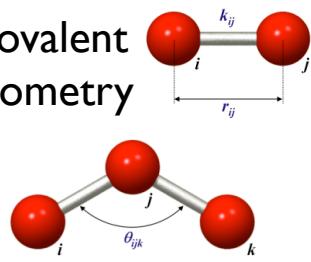


...after model idealization

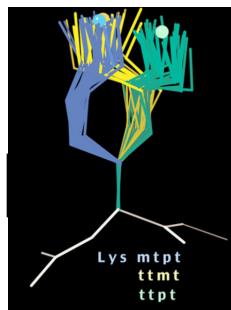


# Special restraints for low-resolution

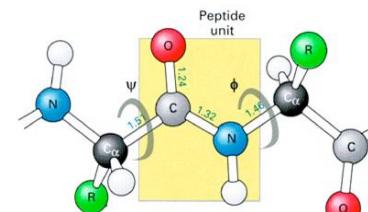
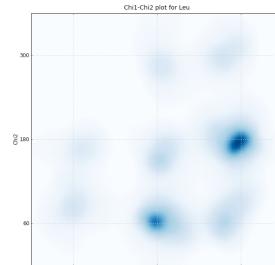
Covalent geometry



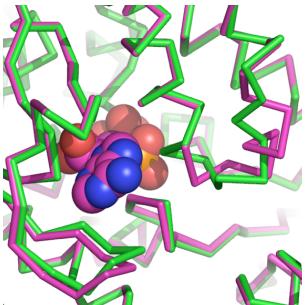
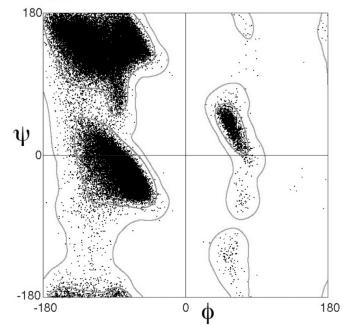
Images from PumMa web site (<http://www.pumma.nl>)



Sidechain distributions

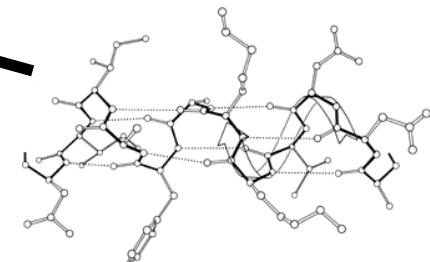


Mainchain distributions



Related structures

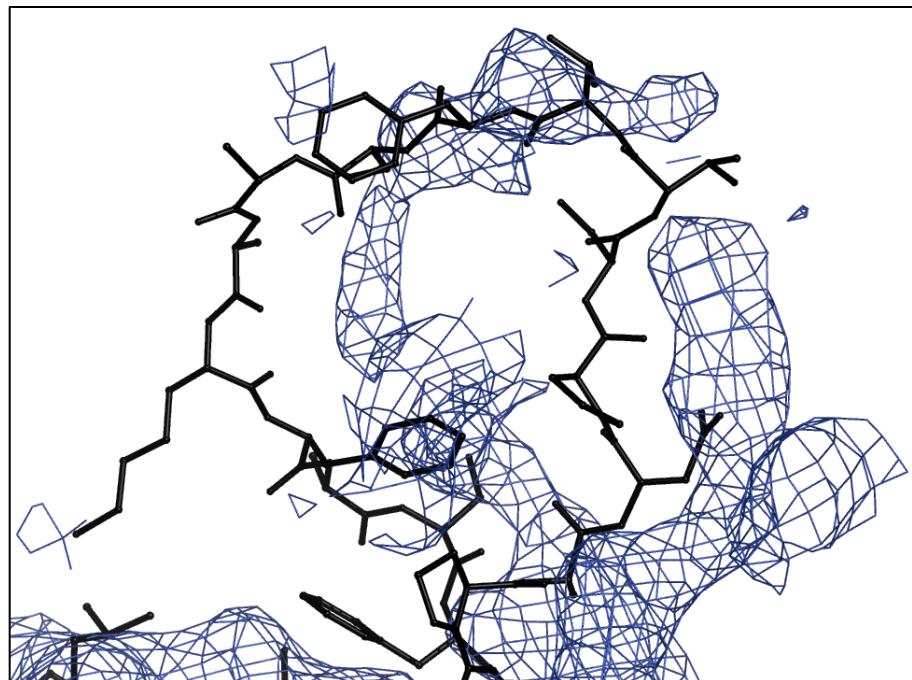
Internal symmetry



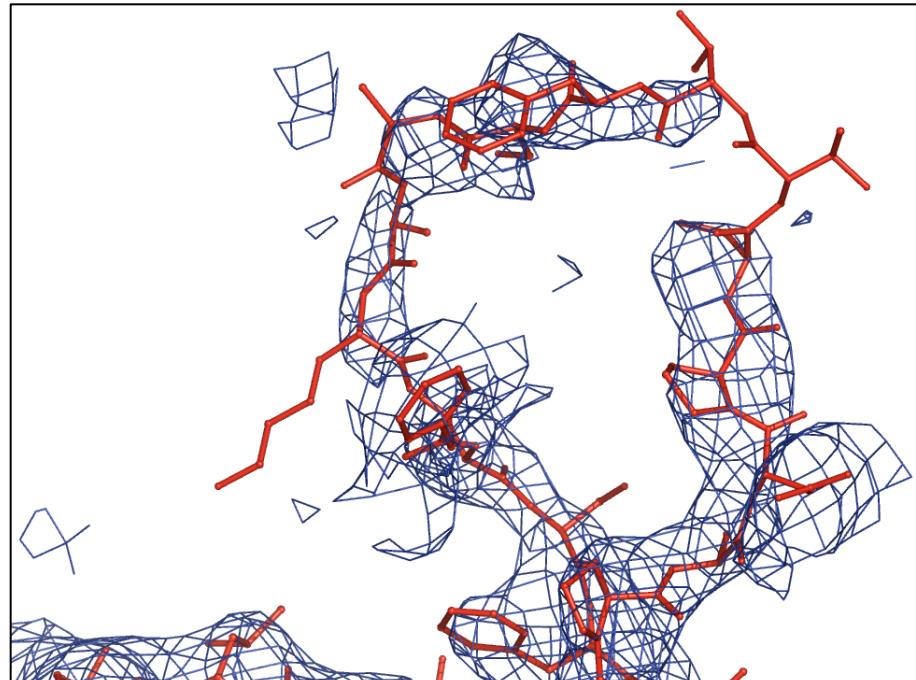
Secondary structure

# Phenix.real\_space\_refine – performance

Start model before refinement

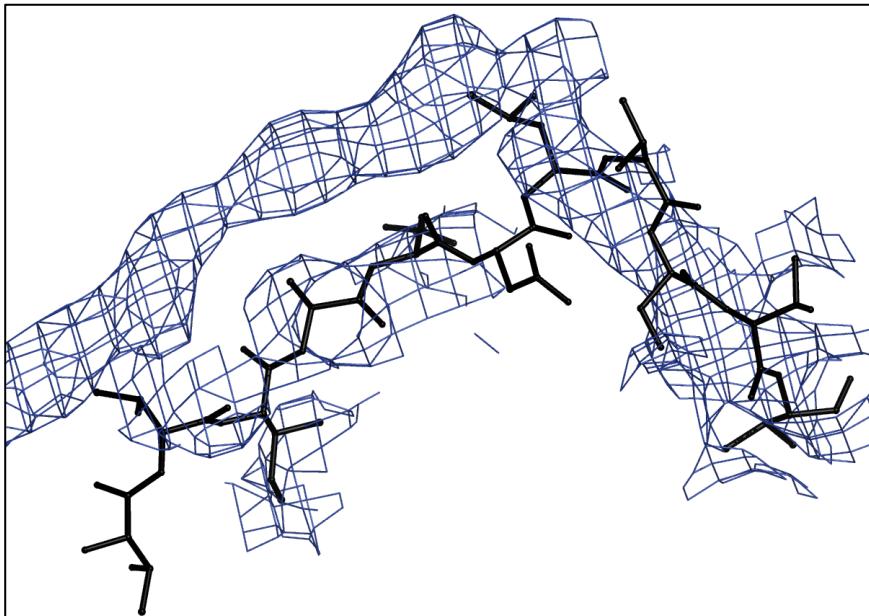


After *phenix.real\_space\_refine*

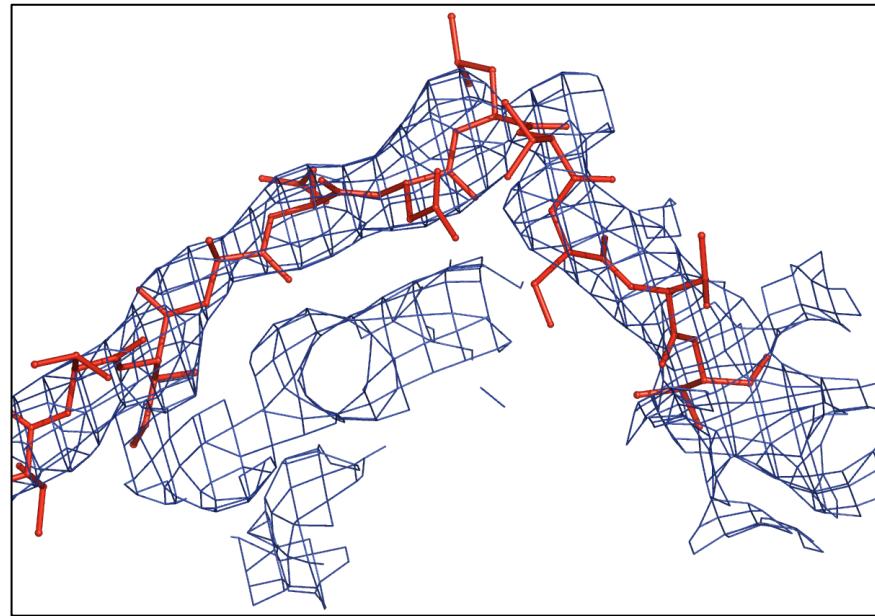


# Phenix.real\_space\_refine – performance

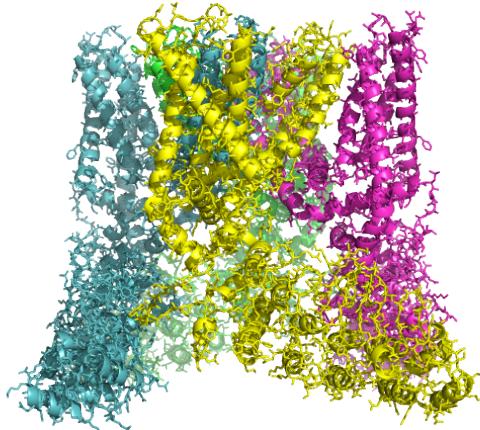
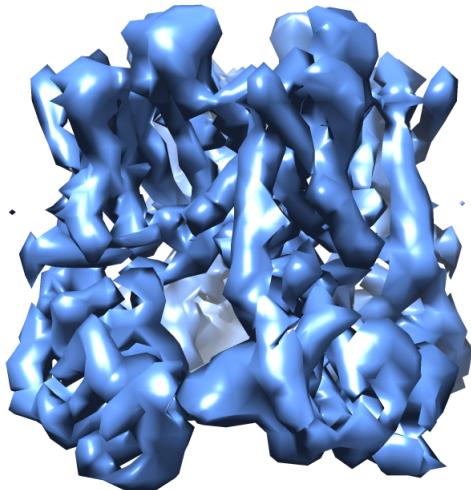
Start model before refinement



After *phenix.real\_space\_refine*



# Examples: 3J5Q, resolution: 3.8 Å

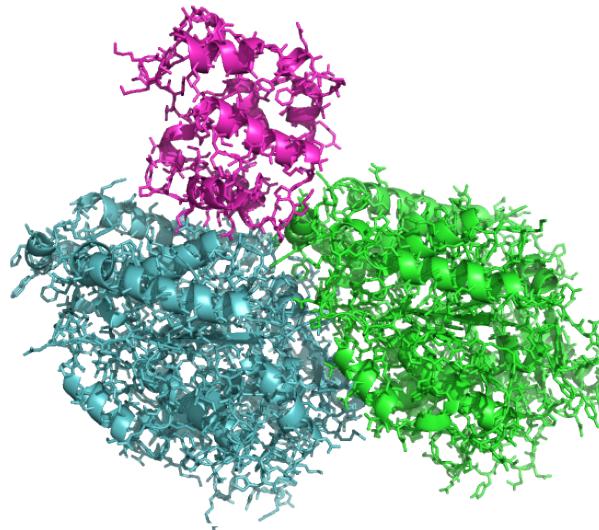
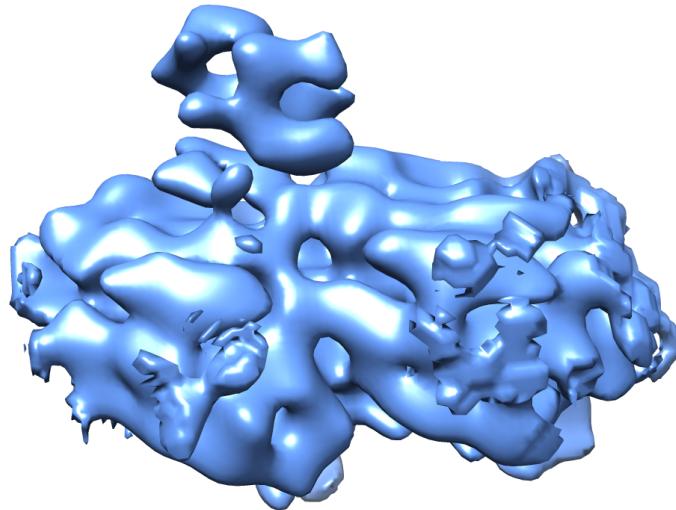


Residues/atoms:  
2,324/17,424

Refinement: 20 min

METRIC	Original	<i>Phenix</i>
Map CC	0.650	0.714
RMSD (bonds/angles)	0.01/1.34	0.01/1.31
Clashscore	100.9	32.84
Rama. outl., %	0.52	0
Rotamer outl., %	27.99	0
C-beta deviations	0	0

# Examples: 3J6P, resolution: 8.2 Å



Residues/atoms:  
949/7,501

Refinement: 15 min

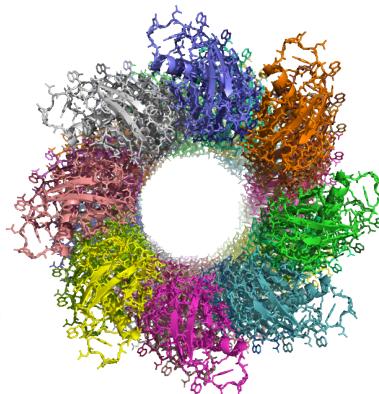
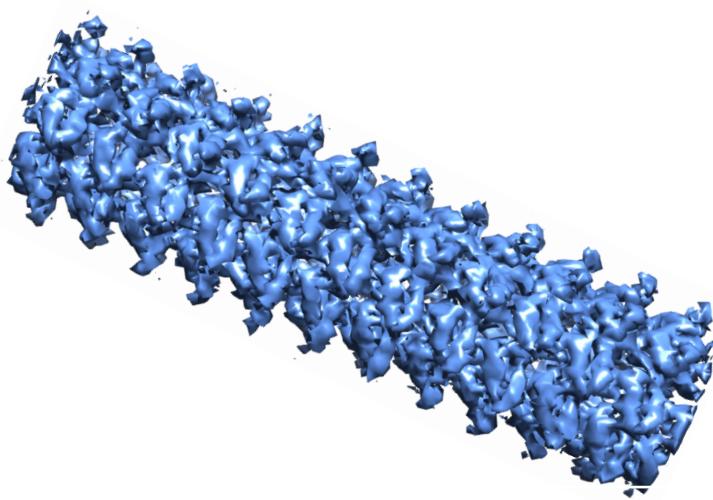
## METRIC

## Original

## *Phenix*

Map CC	0.596	0.743
RMSD (bonds/angles)	0.03/2.34	0.00/1.11
Clashscore	92.37	34.73
Rama. outl., %	2.03	0.54
Rotamer outl., %	26.21	0
C-beta deviations	2	0

# Examples: 3ZEE, resolution: 6.1 Å

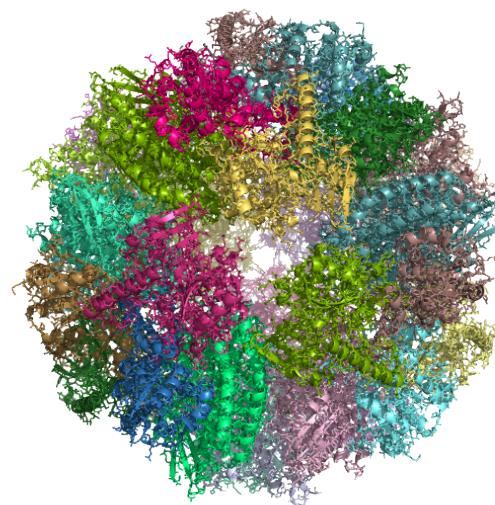
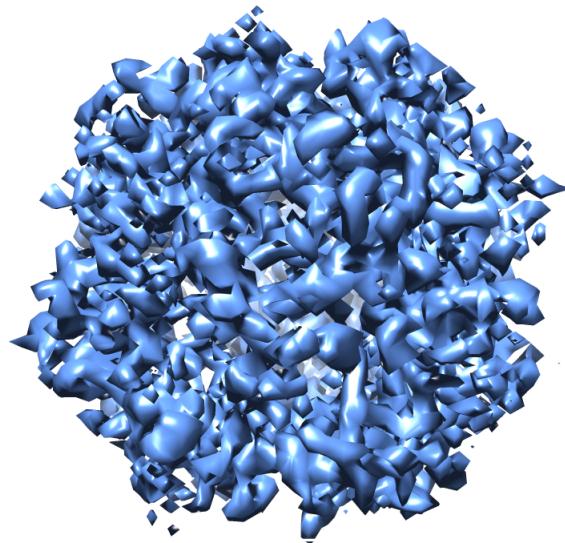


Residues/atoms:  
4,116/32,830

Refinement: 45 min

METRIC	Original	<i>Phenix</i>
Map CC	0.709	0.647
RMSD (bonds/angles)	0.04/4.05	0.01/1.23
Clashscore	18.34	18.59
Rama. outl., %	3.66	0
Rotamer outl., %	24.64	0
C-beta deviations	637	0

# Examples: 4C10, resolution: 3.36 Å



Residues/atoms:  
**10,716/82,404**

Refinement: 173 min

METRIC	Original	<i>Phenix</i>
Map CC	0.645	0.783
RMSD (bonds/angles)	0.02/2.05	0.01/1.21
Clashscore	117.1	18.79
Rama. outl., %	0.11	0.11
Rotamer outl., %	35.51	0
C-beta deviations	24	0

# Resources

PHENIX Documentation [phenix-online.org/version\\_docs/1.9-1692/](http://phenix-online.org/version_docs/1.9-1692/)

## Phenix

*Python-based Hierarchical ENvironment for Integrated Xtallography*

### Phenix Documentation - version 1.9-1692

[Phenix programs and their functions](#)  
[The Phenix graphical interface](#)  
[Dictionary of crystallographic and other terms](#)  
[FAQs: Frequently asked questions](#)  
[How to install, setup and run Phenix](#)  
[Complete Phenix reference documentation](#)  
[Bibliography](#)  
[Index](#)

#### Crystallographic Structure Solution with Phenix

```
graph LR; A[Assess Data Quality] --> B[Experimental Phasing]; A --> C[Molecular Replacement]; B --> D[Model (Re)building]; C --> D; D --> E[Refinement & Validation]; E --> F[Ligand Fitting]; E --> G[Structure Deposition]; F --> G;
```

The flowchart illustrates the Phenix workflow for crystallographic structure solution. It begins with 'Assess Data Quality', which branches into 'Experimental Phasing' and 'Molecular Replacement'. Both of these lead to 'Model (Re)building'. From 'Model (Re)building', the process continues to 'Refinement & Validation', which then leads to 'Ligand Fitting' and finally 'Structure Deposition'.

#### Phenix Documentation for X-ray Crystallography

[Checking data quality](#) | [Experimental phasing](#) | [Molecular replacement](#) | [Model building](#) | [Structure refinement](#)  
[Structure validation](#) | [Ligand fitting](#) | [Making geometry restraints](#) | [Structure deposition](#)

#### Phenix Documentation for Neutron Crystallography

[Structure refinement](#) | [Structure validation](#) | [Making geometry restraints](#) | [Structure deposition](#)

#### Phenix Documentation for Electron Microscopy (EM)

[Structure refinement](#) | [Convert map to structure factors](#) | [Extract box with map and model](#)

# *Phenix* Refinement Tools: Summary

- Real-space refinement: `phenix.real_space_refine`
  - Primarily for Cryo-EM
  - May also be used in low-resolution crystallography
- Reciprocal-space refinement: `phenix.refine`
  - Primarily for crystallography
  - May also be used for Cryo-EM

# Phenix Refinement Tools: Summary

PHENIX home

Citations Reload last job

Coot PyMOL KiNG Other tools Ask for help

**Actions Job history**

**Projects**

Show group: All groups Manage...

Select Delete New project Settings

ID	Last modified	# of jobs	R-free
ringer	Sep 07 2016 05:37 ...	2	---
tmp2	Sep 07 2016 05:23 ...	1	---
5gnn	Sep 07 2016 08:42 ...	1	---
debug1	Sep 05 2016 10:51 ...	2	0.0086
tmp4	Aug 18 2016 07:23 ...	2	---
testing	Aug 11 2016 01:54 ...	1	---
mich	Jul 29 2016 12:47 ...	1	---
almu	Jul 28 2016 10:58 ...	1	---
rchen	Jul 22 2016 11:10 ...	1	---
milya	Jul 15 2016 12:36 ...	2	---
SEM	Jul 14 2016 05:20 ...	1	0.1570

**Data analysis**

**Experimental phasing**

**Molecular replacement**

**Model building**

**Refinement**

**phenix.refine**  
Automated X-ray and/or neutron refinement

**Real-space refinement**  
Automated real-space refinement

**Neutron refinement [alpha]**  
Alternate phenix.refine interface customized for neutron refinement

**DEN refinement [alpha]**  
Deformable elastic network refinement using sparse resolution and molecular replacement

Current directory: /Users/pafonine/Desktop/work/tmp Browse...

PHENIX version dev-svn-000 Project: 5gnn

# Some documented applications using described tools

Bacteriophage epsilon15	Baker et al	2013	PNAS
Brome mosaic virus	Wang et al	2014	Nature
E.coli ribosome	Fischer et al	2015	Nature
Rabbit ryanodine receptor	Yan et al	2015	Nature
Human laforin	Sankhala et al	2015	JBC
Measles virus	Gutsche et al	2015	Science
$\beta$ -galactosidase	Bartesaghi et al	2015	Science
Eukaryotic MCM complex	Li et al	2015	Nature
Viral ATPase	Yu et al	2015	eLife
Anthrax protective antigen	Jiang et al	2015	Nature

# Problems, feedback, questions

- **Talk to us if need help or to report an issue!**

**phenixbb@phenix-online.org**

**bugs@phenix-online.org**

**help@phenix-online.org**

- **Reporting a problem, bug or asking for help:**

- **We can't help you if you don't help us to understand your problem**
- **Send us:**
  - 1) PHENIX version;**
  - 2) Command and parameters used;**
  - 3) Input and output files;**
  - 4) Clearly explain what the problem/question is.**

**Subscribe to PHENIX mailing list: [www.phenix-online.org](http://www.phenix-online.org)**

# Thanks!

---

- ***Phenix* developers and users**
- **Collaborators and colleagues outside of *Phenix* team**
- **Financial support of NIH, LBNL and *Phenix* Industrial Consortium**